Journal of
Experimental Psychology

VOL. 30, NO. 4 APRIL, 1942

AN EXPERIMENTAL STUDY OF THE PHENOMENON OF CLOSURE AS A THRESHOLD FUNCTION [1]

BY JOSEPH M. BOBBITT

Michigan State College

I. THE CONCEPT OF CLOSURE

The concept of closure is one which was originally developed by members of the Gestalt school. However, the term is one which has been used rather widely both by Gestalt writers and by those outside this group; and the meaning of the term has not been invariable. In his 1922 article Koffka (**10**) makes reference to the concept of closure in the following manner: "A good figure is always a 'closed' figure, which the boundary line has the function of closing" (p. 557). Wertheimer (**22**) a year later refers to closure as a tendency for closed organization of stimuli to be preferred to open organizations. Later writers have, however, extended these early and relatively simple applications of the term. Helson (**8**), in his review of Gestalt psychology, relates closure to the reasoning process when he says, "The concept of closure can be applied to thinking. Questions demanding thought arouse configurational processes which are incomplete and which call for closure" (p. 54). Reiser (**18**) has also related closure and reasoning in the following words: "A problem presents itself as an open Gestalt which 'yearns' for solution, and it is the function of thought to find the solution by transforming the open Gestalt into a closed one" (p. 361). He adds, "Such 'closure' or insight reasoning is really reasoning by analogy . . ." (p. 362). Ogden (**16**) in a reply to a query from G. W. Hartmann (**7**) relates closure and insight by saying that all insight implies closure but that closure does not always involve insight. Ogden (**17**) also presents a definition of perception in terms of closure: "Therefore, instead of defining percep-

[1] From the Psychological Laboratory at Northwestern University. The writer wishes to express his appreciation of the help of Dr. A. R. Gilliland who directed this study.

tion as a complex of sensory experience and associated imagery, we have defined it more broadly as the experience of a *closed circuit* of events. Indeed it is because of this 'closure' that we give heed to a series of events and thus *perceive* it" (pp. 126–127). Koffka (11, p. 151) gives closure systematic orientation by identifying it as a special case of the law of Pragnanz. In other words, Koffka considers closure as a manifestation of the tendency of the energies involved in perception to resolve toward a state of equilibrium. If the purpose of the above review were critical rather than expository, one might well express misgivings over the general and loose uses to which the term has been put.

The foregoing review of the uses of the term does not make a definition of closure simple. However, it may be said that the term has been applied to any psychological phenomenon in which a condition of incompleteness, either in the stimulus field or in some phase of the organism's activities, creates a tendency to overcome this incompleteness by perceptual reorganization, by a combination of two or more different experiences, or by some overt activity of the organism. As applied to perception, the term may be defined as a tendency on the part of the organism perceptually to complete stimulus presentations which are physically incomplete. This statement does not mean that the missing parts of, say, a geometrical form, are actually seen, but merely that the parts are organized as belonging to a complete figure and that they represent, as such, a more nearly stable organization than they would represent if separately organized.

II. History

There has been very little experimental work directly on the problem of closure. In 1929 Rosenbloom (**19**) demonstrated a relationship between the tactual two point space threshold and configurational influences in tactual perception. Street (**20**) in 1931 published a Gestalt completion test designed to measure the ability to perceive incomplete pictures. Tiernan (**21**) has recently measured the relative recall and recognition value of open and closed forms in the case of children 7 to 14 years old. In addition to these studies there are many other experiments, some of them of the pre-Gestalt era, which are relevant. In the field of visual perception, the studies of Hempstead (**9**), Meakin (**15**), Zigler (**23**), and Liebmann(**13**) are notable. The work of Gibson (**4**) on the progressive changes in the reproduction of visually perceived forms is also significant. Zigler and Northup (**25**) demonstrated what may well be called closure effects in the case of tactual perception. Finally, the work of Fuchs (**3**) with completion effects in the case of hemianoptic patients cannot be ignored in any consideration of closure.

III. Formulation of Problem

The purpose of the present investigation was that of systematically studying closure in its most simple realm of phenomenal expression, the field of form perception. Specifically, the investigation concerned itself with the question of the perception of physically incomplete triangles, and answers were sought to the following definite questions:

1. Is it possible in the case of triangles, psychophysically to demonstrate a threshold of closure, defined in terms of the percentage of the perimeter of the figure necessary for the occurrence of the phenomenon?

2. Is the threshold of closure in the case of triangles, partially a function of the angular characteristics of the figures?

3. Is the threshold of closure, in the case of triangles, partially a function of the absolute size of the figures?

4. Is the tendency toward perceptual completion of physically incomplete presentations to be explained primarily in terms of the experience of the subject with the complete figure and its various incomplete stages?

5. Is the threshold a stable one, dependent in main part upon a relatively permanent relationship between the organism and the stimulus, or is it primarily determined by the subject's attitude of the movement?

6. Does the measurement of the closure phenomenon give any indication concerning its psychological nature and of the processes underlying it?

IV. The Experiment

A. The Stimulus Forms

1. *General Description and Linear Relationships of the Stimulus Forms.*—Each of the stimulus forms, drawn in black India ink on white paper, consisted of two angles representing incomplete portions of the opposite sides of a triangle (Figs. 1 and 2). The degree of completeness of any one figure was determined, of course, by the length of the lines forming these opposed angles. There were four separate series of triangles. First, there was an equilateral triangle series, which will be designated hereafter as series T-I (Figs. 1 and 2). The dimensions of the complete form of this series were $3\frac{1}{2}'' \times 3\frac{1}{2}'' \times 3\frac{1}{2}''$. Second, there was a large apex isosceles triangle series, which will be indicated hereafter as series T-II. The dimensions of the complete form of this series was $3\frac{1}{2}'' \times 3\frac{1}{2}'' \times 4\frac{2}{3}''$. Third, there was a small apex isosceles triangle series, which will be designated hereafter as series T-III. The dimensions of the complete form of this series were $3\frac{1}{2}'' \times 3\frac{1}{2}'' \times 2\frac{1}{3}''$. Fourth, there was a large apex isosceles triangle series, the angular characteristics of whose forms were the same as those of series T-II and the area and perimeter of whose complete form were equal to those of the complete form of series T-III. The dimensions of the complete form of this fourth series were $2\frac{4}{5}'' \times 2\frac{4}{5}'' \times 3\frac{1}{5}''$. This series will be indicated hereafter as series T-IIR. All lines were $\frac{3}{64}''$ wide. Attention is called to the fact that the base lines in series T-II were twice as long as the base lines in series T-III. Furthermore, the difference between the lengths of the base lines of series T-I ($3\frac{1}{2}''$) and those of series T-II ($4\frac{2}{3}''$) was $1\frac{1}{6}''$, the same as the difference between the length of the base lines of series T-I ($3\frac{1}{2}''$) and those of series T-III ($2\frac{1}{3}''$).

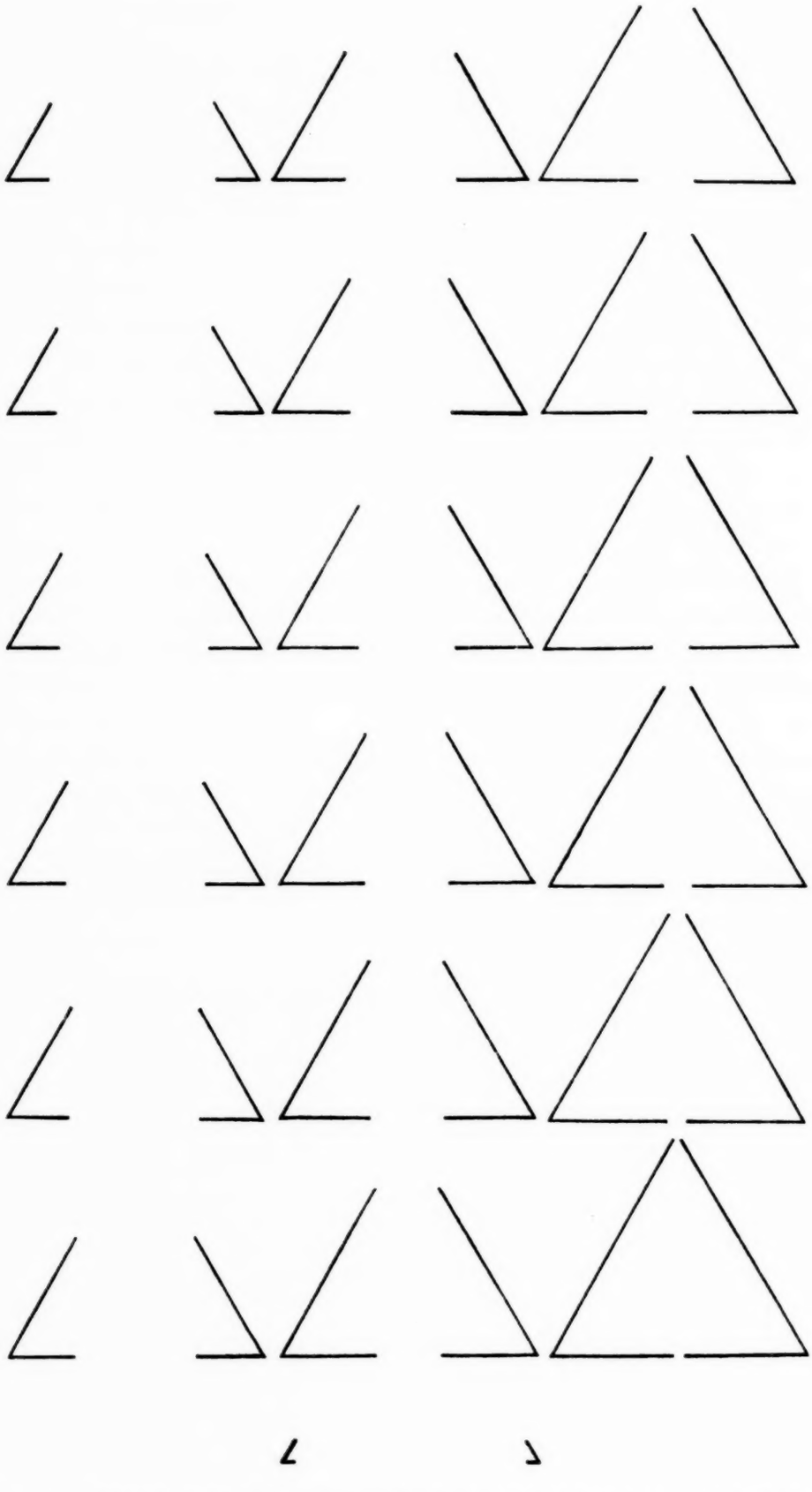


FIG. 1. Triangle series T-I. Form 3 (at bottom) and forms 10 to 27 inclusive. (Photographic reduction.)

A complete set of each series of triangles consisted of 28 forms, the most incomplete of which represented only 3.57 percent of the perimeter of the entire figure and was designated as number 1 in the series. The opposite extreme of the series was, of course, a complete figure and was designated as number 28 in the series. Actually, only form 3 and forms 10 to 27 inclusive for each series was used.[2] This method of numbering the forms makes it possible to identify any form by indicating the series to which it belongs and its place in the series from 1 to 28. Thus, the form represented in Fig. 2 may be indicated as T-I-21.

The way in which the individual forms were constructed deserves detailed description. For the sake of clearness, reference is made to Fig. 2. This figure represents form T-I-21, and there is 75 percent of the total perimeter of the figure present. In this figure, the distance AB represents 75 percent of the side of the triangle of which it is a part. The same statement is true of the distance FE. The distance BC represents 75 percent of half of the base line, and the distance

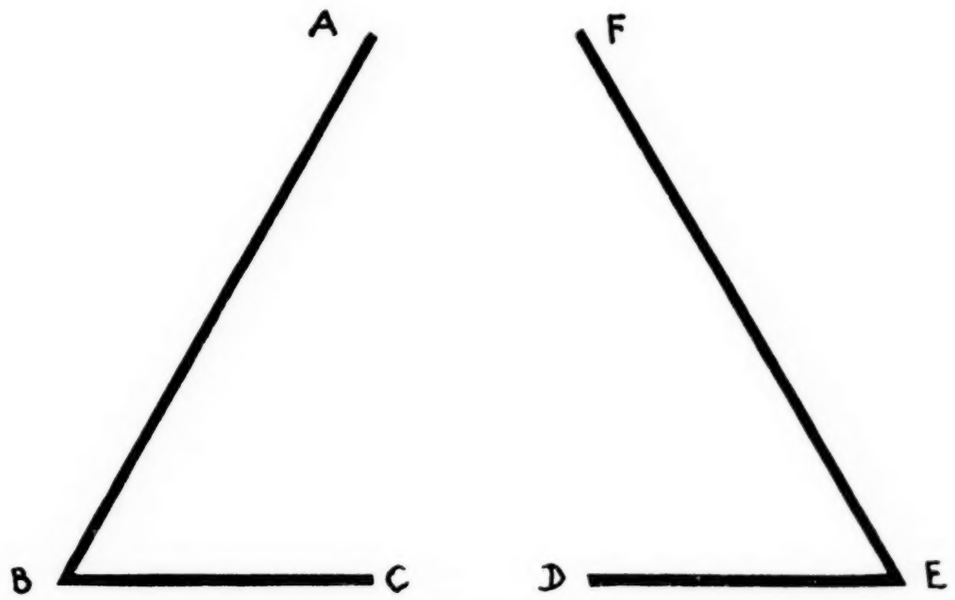


FIG. 2. Form T-I-21. (Actual size.)

DE bears a similar relationship to the other half of the base line. Hence, there is the same percentage of each side of the triangle present as there is of the whole perimeter of the figure. It should be obvious from the foregoing description that forms belonging to different series but having the same location in the series from 1 to 28 were matched in percentage of perimeter both for the whole figure and for each of the parts of the form.

2. *The Areal Relationships of the Stimulus Forms.*—The way in which the percentage of perimeter present in equivalent forms of the four triangle series was matched has just been described. It is significant that the percentage of area represented in equivalent forms was also matched. Referring to Fig. 2, let the percentage which the combined areas of triangles ABC and FDE (each having one imaginary side) represent of the area of the complete triangle be called r_2. Further, let the percentage which is present of the total perimeter of the figure be represented as r_1. It can be shown mathematically that

$$r_2 = r_1^2.$$

Hence, the relationship is independent of the angular characteristics of the triangle. Therefore, the equivalent forms of the four triangle series were matched in terms of area represented as well as in terms of the percentage of the perimeter present. It should be noted, however, that this method of determining the percentage of area represented in a given figure is purely arbitrary and that, from the standpoint of the subject's phenomenological experience, the actual amount of area indicated in a given form of the series might not correspond to the amount represented according to any arbitrary way of determining the percentage present.

[2] Form 3 was not drawn for series T-IIR.

B. Apparatus

The apparatus used in this experiment was a modification of the mirror tachistoscope devised by Dodge (2). The pre-exposure and exposure fields were carefully equated for brightness. The fixation point of the pre-exposure field corresponded in position to the center of the figure on the exposure field. The exposure interval of a tenth of a second was controlled by means of a Bergström chronoscope.

C. Subjects

Twenty-five subjects were used in the investigation. Of these observers, twenty-two were undergraduates, chiefly college freshmen and sophomores; one was a first year graduate student in psychology; one was a second year graduate student in psychology; and one was a trained psychologist. For the purpose of this experiment, all of the undergraduate subjects may be considered as naïve. None of the other three subjects had complete knowledge of the nature of the experiment, but each of them knew the general character of the study.

D. Procedure

Before describing the specific procedure, it is necessary to describe the kind of discrimination that was required of the subjects. The more incomplete forms of each of the series of triangles had what may be termed a quality of *twoness*. That is, the two angles did not appear as necessarily related in any one definite way. They could easily be seen as parts of a triangle, as parts of two separate complex figures, as parts of two triangles, as parts of a trapezoid, or as related in any one of several other ways. In short, while the two angles were recognized as related, they were separately perceived; they did not form one figure. On the other hand, the more complete forms of each series were characterized by a quality of *oneness*. That is, the two angles, while discriminably separate, were spontaneously perceived as being closely related, as belonging to one figure. Obviously, since the forms at one extreme of the series had the quality of oneness and and since the forms at the other extreme of the series had the quality of twoness, at some point in the series of intermediate forms there was a transition point between the forms having predominantly the latter quality and those having predominantly the former quality.

Using a variation of the method of minimal changes, the threshold for closure (defined as the point in the series representing the transition between the forms having the quality of twoness and those having the quality of oneness) was separately determined for each subject for series T-I, T-II, and T-III. The threshold was always stated in terms of the percentage of the total perimeter of the figure present at the point in the series representing the transition between the forms having the quality of twoness and those having the quality of oneness. There are only six different orders in which these series of triangles just mentioned may be presented. One of the orders of presentation was used with five subjects; the other five orders of presentation were used with four subjects each. Hence, each series preceded each other series and was followed by each other series an equal number of times. Therefore, any influence which one triangle series might have on the other series is cancelled in the average results for the whole group.[3] When series T-IIR was used, it was always presented last. Except for the first four subjects, the presentation of the different series of triangles was separated in each instance by at least one day. This provision tended to minimize the influence of one series on the other two series. In the case of eleven subjects, the regular set of tests was followed with the presentation of series T-IIR. Since series T-IIR was given for the purpose of comparing the results obtained by its use with those obtained with series T-II, no subjects were used in this case for whom the first series seen was series T-II. This precaution was taken upon the assumption that the threshold determination for the first series with which a subject was tested might not be quite as reliable as the thresholds obtained with the succeeding series.

The instructions which were read to the subject before presenting the first test series offer the best detailed description of the procedure followed in the actual determination of a threshold.

[3] The fact that one of the orders of presentation was used one more time than were the other orders makes this statement slightly inaccurate, but this slight lack of balance does not defeat the purpose for rotating the presentation orders.

Instructions to Subjects

Here are two incomplete figures (Forms 3 and 27). The one on your right is nearly a complete figure; while the one on your left is very incomplete. Do these two figures look markedly different to you? Would you agree to the statement that the two angles on the right appear to belong together as parts of the same figure, while those on the left do not appear to be closely related to each other in any one definite way; that, in other words, the right-hand angles have a quality of oneness; while the left-hand angles have a quality of twoness?

Obviously, if I should slowly increase the length of the lines forming the left-hand angles by adding small amounts to them, they would at some point lose their quality of twoness and acquire the quality of oneness possessed by the right-hand angles. Here is just such a set of forms, which will be presented in this exposure apparatus. In some of the series, I shall start with a relatively incomplete form and present successive forms which are relatively more complete than the first one. In these cases, your job will be that of determining at what point in the series the angles lose their quality of twoness and acquire the quality of oneness. The series ends at this point. In the other series, I shall start with a relatively complete form and present successive forms which are relatively less complete than the first one. In these cases, your job will be that of determining at what point in the series the angles lose their quality of oneness and acquire the quality of twoness. The series ends at this point. These ascending and descending series will be alternated. These two forms (Forms 10 and 27) represent the most complete and least complete figures that will be used. Remember, however, that, while I shall always start with a form that is relatively complete or one that is relatively incomplete, I shall not always start at one of the extremes of the series. That is, I shall sometimes start with a form that is one or more steps removed from these two forms which represent the extremes of the series that will be used.

In order to give you a more clear idea of what the transition from oneness to twoness is, I am going to show you the whole series outside of the exposure device. Indicate the point at which any decided difference in the angles becomes apparent to you. (After this one preliminary trial, subject was placed before tachistoscope.)

Preliminary to the real experiment, I shall show you the same card two or three times in order that you may become adapted to the apparatus. You should sit in a comfortable position and look through this eyepiece. You will see a rectangular white field, upon which there are two gray circles which overlap, forming a small, vertical, black oval in the center of the field. Adjust your eyes so that the oval is as nearly as possible in the center of the rectangle. When you have made this adjustment, let me know. Before each presentation, I shall give you a ready signal, and very shortly thereafter I shall expose the figure. These trials will be the preliminary ones of which I spoke.

Now we are ready for the real experiment. After each presentation, reply as quickly as possible, that is, without deliberation, with the word *two* if the angles have the quality of twoness; with the word *one* if the angles have the quality of oneness. Please remember that there are no right or wrong answers. This experiment is not in any sense a test of your ability.

There will be rest periods at appropriate intervals. Have you any questions?

It is true that the instructions could be criticized due to the fact that they make suggestions concerning the *oneness* and *twoness* qualities. The answer to this charge is the fact that preliminary work showed these qualities to exist. The suggestions rather than biasing the subjects merely made clear the basis of discrimination which was involved. Also, any such suggestions would not affect the location of the threshold. Further, this influence would not explain the similarity of the thresholds for the different subjects or other characteristics of the data secured.

Attention should be called to several additional points. The parenthetical material in the instructions is explanatory and was not, of course, read to the subject. Form 3 was used as a basis for establishing the subject's criterion of *twoness;* form 27, for establishing his criterion of *oneness*. Form 3, of course, was never used in the actual threshold determination series. A complete test for one triangle involved twenty separate threshold determinations, consisting of ten ascending series and of ten descending series, arranged in the counterbalanced order for the purpose of equalizing as much as possible such factors as fatigue and practice. With thirteen of the twenty-five subjects the order of ascending and descending series was adada dadad dadad

adada; with the other twelve subjects the order of presentation was dadad adada adada dadad. This arrangement prevented the average results for all subjects from being biased as a result of the influence of the initial presentation (either ascending or descending) upon the results of subsequent presentations. The first two presentations of the stimuli included all eighteen forms (forms 10 to 27 in the series). On the basis of the results of these two presentations, a series of fourteen forms was chosen so that the subject's threshold was approximately at the middle of the series. After the range of forms was chosen for a subject, the same range was used for all the series of triangles with which he was tested. All exposure times were a tenth of a second.

Presentation of stimuli were made at the average rate of three or four forms a minute, including the time used for rest periods and for recording the data. Hence, the determination of a subject's threshold for one triangle series, involving between 140 and 150 stimulus presentations, took between forty and fifty minutes. The time necessary to record the data and to rearrange the stimulus cards after each series of ascending or descending presentation of stimuli provided short rest periods. A rest interval of three or four minutes was given upon the completion of the first half of the test.

V. Results

A. Results for Individual Subjects

As mentioned in the previous chapter, the results for any one subject consisted of twenty separate determinations of the threshold of closure for each of the triangle series T-I, T-II, and T-III. For

TABLE I

Results of Individual Subjects for Series T-I and T-II

Subject	Mean Closure Threshold in Percent of Perimeter		S.D. of Distribution in Percent of Perimeter	
	T-I	T-II	T-I	T-II
1	77.86±1.25*	85.00±0.46	5.43±0.86	2.00±0.32
2	73.03±1.07	71.43±1.04	4.71±0.75	4.57±0.71
3	74.82±1.04	77.32±0.86	4.57±0.71	3.79±0.60
4	64.46±1.14	68.21±1.07	5.00±0.79	4.71±0.75
5	75.71±0.68	85.00±0.54	3.00±0.46	2.29±0.36
6	59.46±0.61	63.21±0.50	2.61±0.43	2.14±0.32
7	63.57±1.39	64.11±0.93	6.00±0.96	4.00±0.64
8	62.68±0.68	66.25±0.54	2.89±0.46	2.39±0.39
9	80.71±0.50	82.86±0.54	2.21±0.36	2.29±0.36
10	74.29±0.79	78.39±0.54	3.39±0.54	2.39±0.39
11	69.46±0.75	79.11±0.46	3.21±0.50	2.04±0.32
12	71.79±0.54	76.79±0.68	2.36±0.36	3.00±0.46
13	55.18±1.07	54.82±0.54	4.71±0.75	2.36±0.36
14	78.57±0.39	84.64±0.32	1.79±0.29	1.43±0.21
15	68.57±0.75	67.14±0.82	3.21±0.50	3.57±0.57
16	84.46±0.36	83.57±0.89	1.50±0.25	3.89±0.61
17	88.57±0.39	90.54±0.46	1.64±0.25	2.04±0.32
18	73.57±0.64	73.93±1.14	2.75±0.43	5.00±0.79
19	60.54±0.82	63.75±1.00	3.64±0.57	4.29±0.68
20	74.64±0.61	81.61±0.61	2.61±0.43	2.61±0.43
21	66.07±0.57	73.75±0.46	2.43±0.39	2.04±0.32
22	76.43±1.00	75.18±1.21	4.36±0.68	5.25±0.82
23	83.21±0.61	83.39±0.54	2.68±0.43	2.36±0.36
24	65.00±0.79	68.57±0.64	3.39±0.54	2.79±0.43
25	76.61±0.75	78.04±0.68	3.29±0.54	3.04±0.46

* In this table and all succeeding ones, variability measures of means and of standard deviations are in terms of the standard error.

TABLE I (*Continued*)

RESULTS OF INDIVIDUAL SUBJECTS FOR SERIES T-III AND T-IIR

Subject	Mean Closure Threshold in Percent of Perimeter		S.D. of Distribution in Percent of Perimeter	
	T-III	T-IIR	T-III	T-IIR
1	78.39±0.39		1.79±0.28	
2	74.64±0.64		2.86±0.46	
3	75.89±0.86	76.43±0.64	3.71±0.57	2.75±0.43
4	63.57±1.18		5.07±0.79	
5	70.18±0.54	83.39±0.39	2.36±0.36	1.71±0.29
6	57.68±0.54		2.36±0.36	
7	60.18±0.68	65.71±1.36	3.03±0.46	5.96±0.93
8	56.25±0.86	67.50±0.46	3.71±0.57	2.07±0.32
9	80.71±0.68		2.96±0.46	
10	71.25±0.93		4.00±0.64	
11	61.61±0.64	75.89±0.43	2.75±0.43	1.93±0.32
12	62.50±0.64	75.71±0.39	2.86±0.46	1.64±0.25
13	47.50±0.68		2.89±0.46	
14	77.14±0.43	77.14±0.43	1.93±0.32	1.93±0.32
15	55.89±0.68		3.04±0.46	
16	83.39±0.54		2.36±0.36	
17	77.50±0.79	87.14±0.43	3.50±0.57	1.93±0.32
18	50.71±1.21		5.25±0.82	
19	57.32±0.93	58.21±0.71	4.00±0.64	3.11±0.50
20	74.64±0.75	79.82±0.54	3.29±0.54	2.36±0.36
21	64.64±0.46		2.07±0.32	
22	72.50±1.00		4.32±0.68	
23	77.32±0.39		1.71±0.29	
24	59.11±0.75	62.86±0.64	3.29±0.54	2.75±0.43
25	78.03±0.75		3.25±0.50	

eleven subjects, the procedure also included the use of triangle series T-IIR. For a given determination, the threshold of closure was assumed to fall halfway between the two forms representing the change from the quality of twoness to the quality of oneness. For example, if, on an ascending series in which the forms were becoming progressively more complete, the subject reported a quality of twoness for form 20 in the series and a quality of oneness for form 21, the closure threshold for that determination was assumed to fall at a point in the series halfway between the two forms. It will be remembered that the threshold is always stated in terms of the percentage of the total perimeter of the figure present at this point in the series. Using the twenty separate thresholds thus secured, the mean closure threshold and the standard deviation were determined for each subject. Also, the standard error of each of these measures was computed. Table I summarizes these values for each subject.

B. Results for Entire Group of Subjects

1. *Group Mean Closure Thresholds.*—The group results were based in each case upon the mean thresholds for individual subjects. Using these values, the group's mean closure threshold, the standard devia-

tion and the standard error of each of these measures were computed for each triangle series. Table II summarizes these values for triangle series T-I, T-II, and T-III.

TABLE II

SUMMARY OF GROUP RESULTS
(SERIES T-I, T-II, AND T-III)
$N = 25$

Series	Mean Closure Threshold in Percent of Perimeter	S.D. of Distribution in Percent of Perimeter
T-I	72.00±1.68	8.18±1.14
T-II	75.06±1.75	8.64±1.21
T-III	67.63±2.04	10.00±1.43

2. *Difference Between Group Mean Thresholds for the Three Main Triangle Series.*—The differences between the group's mean closure thresholds for these three triangle series are shown in Table III.

TABLE III

SUMMARY OF DIFFERENCES BETWEEN GROUP MEAN THRESHOLDS
(SERIES T-I, T-II, AND T-III)
$N = 25$

Mean Threshold in Percent of Perimeter	Mean Threshold in Percent of Perimeter	Diff. in Percent of Perimeter	$\sigma_{diff.}$ in Percent of Perimeter	C.R.$_{\sigma}$
Mean, T-II 75.06±1.75	Mean, T-I 72.00±1.68	3.06	0.68	4.50
Mean, T-I 72.00±1.68	Mean, T-III 67.63±2.04	4.37	1.07	4.08
Mean, T-II 75.06±1.75	Mean, T-III 67.63±2.04	7.43	1.18	6.30

The standard error of the difference and the critical ratio are also given in each case.

3. *Inter-Correlations of the Three Main Triangles Series and the Reliabilities of the Separate Series.*—Table IV shows the three possible coefficients of correlation between the performance of the subjects on the three triangle series, T-I, T-II, and T-III. The coefficients were computed by the method of rank differences. The correlations indicate the degree to which a subject's mean threshold tended to occupy on each of the three series of triangles the same relative position with regard to the distribution of thresholds of the various members of the group. Also, these correlations between different triangle series indirectly indicate the coefficient of reliability for the

separate triangle series. The highest possible inter-correlation between two variables is equal to the geometric mean of the separate coefficients of reliability of the two variables. Using the lowest correlation presented below (+.816, between series T-II and T-III) as the basis of computations, the following statements may be made. (1) If one of the two series represented be assumed to have a perfect

TABLE IV

RANK ORDER COEFFICIENTS OF CORRELATION (RHO)
(SERIES T-I, T-II, AND T-III)
$N = 25$

	T-I	T-II	T-III
T-I	——	+.908±.023*	+.877±.033
T-II		——	+.816±.047
T-III			——

* Probable errors.

reliability, the other series would have to have a coefficient of reliability of .666. (2) If the reliabilities of the two triangle series are assumed to be equal, each series will have a coefficient of reliability of .816, the same as the coefficient of correlation between the two series. Of the two assumptions, the latter seems much the more reasonable. Hence, it seems justified to suppose that each of the series of triangles represented in this correlation has a reliability coefficient of approximately the same size as the reported coefficient of correlation between the two series (.816). Attention is again called to the fact that these statements are based upon the lowest of the three reported correlations. Split half reliabilities were not computed, because they would not indicate the test-retest reliability of the triangle series, but merely the tendency of the subjects to be consistent during any one experimental session.

4. *First Half-Second Half Analysis of the Data.*—In order to ascertain whether or not there was a tendency for the closure threshold to shift significantly in either direction during the course of the experiment, two separate closure thresholds were determined for each subject for each of the three triangle series, T-I, T-II, and T-III. One of these measures represented the subject's mean threshold on the first half of the twenty separate threshold determinations; the other measure represented the subject's mean threshold in the last half of the test. Using these individual results, the group's mean closure threshold and the standard error of the mean for the first and second halves of the test were separately determined for each

TABLE V

COMPARISON OF GROUP RESULTS FOR FIRST HALF AND SECOND HALF OF THE THRESHOLD DETERMINATIONS (SERIES T-I, T-II, AND T-III)

$N = 25$

	T-I	T-II	T-III
Group Mean Threshold for First 10 Trials in Percent of Perimeter	71.91±1.64	74.57±1.75	67.50±1.96
S.D. of Distrib. for First 10 Trials in Percent of Perimeter	8.04±1.14	8.61±1.21	9.54±1.36
Group Mean Threshold for Second 10 Trials in Percent of Perimeter	72.08±1.71	75.56±1.82	67.76±2.18
S.D. of Distrib. for Second 10 Trials in Percent of Perimeter	8.46±1.21	8.82±1.25	10.61±1.50
Diff. Between Means in Percent of Perimeter	−0.17	−0.99	−0.26
$\sigma_{diff.}$ in Percent of Perimeter	0.46	0.39	0.50
C.R.$_{\sigma}$	0.37	2.54	0.52

of the three series of triangles. Table V summarizes these results for the group. The table also shows the differences between the first and second half group mean thresholds, the standard errors of these differences, and the corresponding critical ratios.

C. Results for the Area Control Group

Attention should be called to the fact that the three triangle series, T-I, T-II, and T-III, differ from each other not only in terms of their angular characteristics but also in terms of their areas and of their perimeters. Of the three series of triangles, the complete form of series T-III had the smalles area (3.85 sq. in.); that of series T-II, the largest area (6.09 sq. in.). In order to determine whether or not part of the significant differences between the mean group thresholds of the three series of triangles was the result of differences in the areas and perimeters, triangle series T-IIR was used with eleven subjects. It will be remembered that the angular characteristics of the forms of this series were the same as those of series T-II and

TABLE VI

SUMMARY OF RESULTS FOR AREA CONTROL GROUP (SERIES T-I, T-II, T-III, AND T-IIR)

$N = 11$

Series	Mean Closure Threshold in Percent of Perimeter	S.D. of Distribution in Percent of Perimeter
T-I	71.40±2.64	7.89±1.68
T-II	76.15±2.93	8.79±1.86
T-III	66.77±2.68	8.07±1.71
T-IIR	73.62±2.82	8.50±1.82

that its complete form had the same area and perimeter as did the complete form of series T-III. Hence, the results obtained with triangle series T-IIR give a basis for determining the influence upon the closure threshold of differences of absolute size. In order to facilitate discussion the eleven subjects with whom series T-IIR was used will be called the area control group.

1. *Group Mean Closure Thresholds.*—Table VI summarizes the area control group results for triangle series T-I, T-II, T-III, and T-IIR. The table indicates for each triangle series, the group's mean closure threshold, the standard deviation, and the standard error of each of these measures.

2. *Differences Between Group Mean Thresholds for the Four Triangle Series.*—The differences between the mean closure thresholds for the group for these four triangle series are shown in Table VII.

TABLE VII

SUMMARY OF DIFFERENCES BETWEEN MEAN THRESHOLDS FOR AREA CONTROL GROUP (SERIES T-I, T-II, T-III, AND T-IIR)

$N = 11$

Mean Threshold in Percent of Perimeter	Mean Threshold in Percent of Perimeter	Diff. in Percent of Perimeter	$\sigma_{diff.}$ in Percent of Perimeter	C.R.$_\sigma$
Mean T-IIR 73.62±2.82	Mean, T-I 71.40±2.64	2.22	1.14	1.95
Mean, T-II 76.15±2.93	Mean, T-IIR 73.62±2.82	2.53	0.93	2.72
Mean, T-IIR 73.62±2.82	Mean, T-III 66.77±2.68	6.85	1.68	4.08
Mean, T-II 76.15±2.93	Mean, T-I 71.40±2.64	4.75	0.93	5.11
Mean, T-I 71.40±2.64	Mean, T-III 66.77±2.68	4.63	1.14	4.06
Mean, T-II 76.15±2.93	Mean, T-III 66.77±2.68	9.38	1.53	6.13

The standard error of the difference and the critical ratio are also given in each case.

VI. Discussion

A. Comparison of the Present Experiment and Traditional Threshold Studies

The present experiment represents the application of the threshold concept to a fairly complex perceptual process. The method followed is strictly analogous to that used in traditional threshold

measurements. It should be noted that the data secured in this study are similar to those secured in measurements of simple sensory thresholds. By reference to Table I it may be seen that the separate threshold determinations for each subject varied within a small range around a very definite central tendency. Table V indicates that there is relatively little tendency for the threshold to shift during the course of the measurement. These observations justify the statement that the present investigation involves the measurement of a true threshold function.

B. The Main Experimental Finding

1. *Analysis of the Main Experimental Finding.*—There are obviously many factors which may influence the closure threshold. They may be legitimately divided, however, into two classes: (1) intrinsic factors, which are those connected with the figure as such; and (2) extrinsic factors, which are those external to the figure itself. In this latter group are, for example, such determinants as exposure time, intensity of illumination, and all other strictly methodological variables. The intrinsic factors, on the other hand, are those dependent upon the internal relationships of the figure and upon the general phenomenological characteristics of the form. In the present study, all extrinsic variables were, as far as possible, held constant, and the only differential influences upon the closure threshold in the various triangle series were the intrinsic ones. In the light of this statement, then, *the main experimental finding of the present study is the fact that there are statistically significant differences between the thresholds of closure of different triangles* (*See Table III*) *which can be explained only in terms of the influence which the intrinsic factors involved have upon the size of the threshold.* The specific nature of these variables in the present case is indicated in the analysis of the two ways in which the three main series of triangles, series T-I, T-II, and T-III, differed from each other. (1) Their angular characteristics were different, which means that the proportional relationships of their sides to each other were also different. (2) Their absolute sizes (areas) were different, which means that their perimeters (specifically, in this case, their base lines) were of different absolute lengths. These two variables are also functions of each other. The observed threshold differences must, then, be explained in terms of the effect of one or of both of these two factors upon the location of the closure threshold.

The data obtained with triangle series T-IIR give some information concerning the influence which these two factors, the angular characteristics of the forms and their area, had on the closure thresh-

old. The forms of this series differed from those of series T-III only in terms of their angular characteristics; that is, these two series were equated in terms of absolute size. Table VII shows that even when this factor is equated, the threshold difference between the two series of triangles does not disappear. This difference, as the table shows, is statistically significant as expressed by a critical ratio of 4.08 in terms of the standard error of the difference. Moreover, series T-IIR and series T-II differed from each other only in terms of their size; that is, these two series were equated in terms of their angular characteristics. In this case, a small measure difference is obtained which fails to meet the conventional standard of statistical significance (See Table VII). However, the critical ratio of 2.72 in terms of the standard error of the difference is large enough to suggest that the difference is possibly a real one and that a greater number of cases would demonstrate its statistical significance.[4] In this connection, it is interesting to note that Fuchs (**3**) found that completion effects on the blind half of the retina in the case of hemianoptic patients are largely but not absolutely independent of the absolute size of the figures used. He also found that in most cases about 75 percent of the figure had to be represented on the good half of the retina in order for completion effects to occur. The group mean closure thresholds obtained in the present study for the three main series of triangles range between 67.63 percent to 75.08 percent of the total perimeter. It is possibly significant that the present writer's results with normal subjects and those obtained by Fuchs with hemianoptic cases are as similar as they are. In recapitulation, then, the main experimental finding of this study shows that factors other than the mere percentage of the figure present influence the threshold of closure. On the basis of the experimental results, two statements concerning these other factors may be made. (1) The data conclusively show that the angular characteristics of the forms are important determinants of the closure threshold. (2) The factor of absolute size appears slightly to modify the threshold of closure, but the measured difference does not meet the conventional test of statistical significance; and there is the further possibility that this observed difference may be a reflection of the fact that the smaller

[4] It is possible that this difference is not really based upon the fact that the absolute sizes of the forms were different, but upon the fact that the forms of series T-IIR had a different relationship to the constant size exposure field than did the similar but larger one, of series T-II. In fact this explanation appears to be the more sophisticated one of the two. This factor of the relationship between the exposure field and the size of the figure, while possibly influencing the differences between the group mean thresholds for the main series of triangles, T-I, T-II, and T-III, does not completely explain them. This last statement is based upon the fact that a highly significant difference (critical ratio, 4.08) was obtained between the group mean thresholds of series T-IIR and T-III, the forms of both of which covered the same proportional part of the exposure field.

forms (those of series T-IIR) had a different relationship to the constant size exposure field than did the larger forms (those of series T-II).

2. *Significance of the Main Experimental Finding. a. The Main Experimental Finding in the Light of Empirical Concepts.*—While there is in the literature no formal consideration of closure by traditional psychologists upon which to base the statement, it appears patent to the writer that an explanation of this phenomenon from the existential, functional, or behavioristic point of view would have to be stated in terms of the influence of experience upon perceptual responses. Specifically, an explanation stated in such terms would have to assume that a person's experience with a complete figure and its various incomplete forms would result in his recognition of the relationship of the incomplete representation to the complete figure. Such a theory would have to assume that after sufficient experience of this kind a person would be able to recognize a partial representation of a figure as belonging to a complete form. The amount of experience would, of course, determine the facility with which the recognition process takes place. Stated in its most simple terms, then, closure would become, in terms of this explanation, a phenomenon dependent upon perceptual habits.

In the writer's opinion, this explanation is unable to account for the fact that there are statistically significant differences between the closure thresholds of different kinds of triangles. In the first place, the existence of the differences necessitates the assumption that one figure has an experience advantage over the other one. It is conceivable that such a difference in the experience with different kinds of triangles might exist for any one subject, but this possibility offers no explanation of why the same triangle enjoys this superiority for nearly all subjects. Incidentally, the equilateral triangle which might plausibly be expected to be the one with which a person normally has the most experience has neither the highest nor the lowest threshold of the three kinds of triangles used in this study. In the second place, it will be remembered that in the case of each triangle series the group mean closure threshold based upon the second half of the separate determinations for each subject was slightly higher than was the threshold based upon the first half of the determinations for each subject (See Table V). None of these differences is, however, statistically significant. Hence, while the subject had some experience with the various incomplete forms of each triangle series during the course of the experiment, it did not significantly influence his closure threshold. Any empirical explanation of closure would, of course, have to assume that this experience would result in a lowering of the threshold. As a matter of fact, the obtained differences between the first and second half mean thresholds indicate that any

changes which did occur during the experiment were in the direction of higher thresholds rather than in the direction of lower ones. There are, of course, two obvious objections to this last criticism of the empirical explanation of closure. First, it is possible that the subjects had already had so much experience with partial representations of triangles that the small amount of experience gained during the experiment was ineffective in causing any difference in the closure threshold. In other words, the process of learning to recognize complete figures upon the basis of perception of their parts had possibly already reached a final level. While this contention is very reasonable, it leaves to be explained the fact that this final level was reached at different points in the case of different triangles. Second, the slight increase in the closure threshold during the experiment might be explained on the basis of fatigue. This claim is not very tenable in view of the fact that only one triangle series was presented during one experimental session (except for four subjects); that the subject had a short rest interval after each six to ten stimulus presentations; that a rest of three or four minutes was given at the end of the first half of the test. In summary, then, an explanation of closure in terms of traditional psychological concepts would have to be stated in empirical terms. It is difficult to offer an interpretation of the results of the present study on the basis of such an explanation because its use in this connection necessitates the assumption that nearly all of the subjects used had had significantly different amounts of experience with the different kinds of triangles. The present writer feels that this assumption is unjustified. Also, the fact that the closure thresholds did not change significantly during the experiment and that what change did occur was in a direction opposite to what one would anticipate on the basis of empirical principles militates against the acceptance of any explanation of closure involving the effects of experience as the primary determinant of its occurrence.

b. The Main Experimental Findings in the Light of Gestalt Concepts.—The Gestalt writers have been no more definite than have the empiricists in giving a detailed explanation of what happens when closure occurs. The present study does not supply this information; but it does demonstrate that progressive changes in the degree of completeness of a figure result in a rather sudden change in the way the presentation is organized. It is possible to suggest the basis of this change in the light of the principles of brain activity supported by the Gestalt writers. These writers speak of perception as being the result of potential differentials in the cortex (**12**, pp. 185–188), and then further suggest that the potential patterns set up internal organizational forces which tend to resolve themselves to a simple stable pattern (**11**, pp. 138–176). Such a hypothesis of brain ac-

tivity during perception may be applied to the present situation. The primary assumption involved in such an explanation is that each part of the figure (say, the two angles of the figures used in the present study) is represented cortically by an area having a potential level different from that of the rest of the cortex. A secondary assumption is that two areas of potential difference will be subject to organizational forces and that this organization will tend to be as simple and well balanced as possible. There are, however, two fundamentally different ways in which the two angles of the forms used here may be organized. That is, they may be organized in a dual manner as two separate figures having a balanced relationship to the whole visual field, or they may be organized in a unitary manner as parts of one figure. Obviously, the relative strength of the forces acting toward dual and unitary structurization of the field is a function of the percentage of the figure present. When the figure is very incomplete the tendency to perceive each angle separately is stronger than is that to see the angles in a unitary manner. With the slow increase of the percentage of the perimeter present, a critical point is reached at which the tendency toward a unitary organization becomes stronger than the tendency toward a dual structuring of the field. Phenomenologically, this point corresponds to the experience of closure. In terms of the present study this point represents the place in the series at which occurred the shift from the quality of twoness to the quality of oneness.

The different thresholds for the different kinds of triangles, from this point of view, merely seems to indicate that the organization forces are affected not only by the completeness of the figure, but also by the kind of figure it is (*i.e.* by the angular characteristics of the forms, in this case). Attention may be called to the fact that there is a perfect correspondence between the size of the indicated apex angle and the threshold of closure. That is, the small apex triangle series (T-III) had the lowest threshold; the large apex triangle series (T-II) had the largest one. There are other experimental data which may be considered consistent with this finding. Zigler (**23**) found that the apex of a visually perceived triangle has the greatest attributive significance. Collier (**1**) found acute angle forms to be easier to recognize in peripheral vision than are obtuse angle forms. Zigler and Barrett (**24**) found acute angle forms (and specifically triangles) are easier to perceive tactually than are obtuse angle forms; and Lindemann (**14**) showed the gamma movement to be correlated with the shape characteristics of the figure. In the case of triangles having horizontal base lines, a violent gamma movement occurred at the apex. Hence, angular forms are shown to be highly organized. Further, acute angle forms appear to have a su-

periority in this respect over obtuse angle forms. Finally, in the case of triangles, the apex appears to be the point of the highest expression of organizational dynamics. In summary, then, two statements may be made. (1) Closure may be considered as an expression of organizational dynamics. (2) The differences between the closure thresholds of the three main series of triangles appears to be largely due to the fact that the tendency toward unitary organization of the field is partially a function of the angular characteristics of the figure and possibly, but to a lesser extent, of the absolute size of the forms. Attention is called to the fact that this explanation is offered as one of several possible interpretations of the present data. It is, at any rate, adequate for the interpretation of the present results, but subsequent research may force alterations in the specific aspects of the explanation.

C. Secondary Experimental Findings

1. *The Significance of the First Half-Second Half Analysis of the Data.*—It has already been noted that the group's mean closure threshold for each trianlge series computed upon the basis of each subject's first ten threshold determinations was not significantly different from the group's mean threshold computed upon the basis of each subject's last ten determinations. It has already been suggested, since fatigue does not appear to be a very likely explanation of the small increase in the threshold during the experiment, that experience within the limits of that given in this experiment does not appear to influence the closure threshold. Further elaboration of this point is not necessary, but the writer wishes to call attention to the similarity between this finding and Gottschaldt's (**5, 6**) demonstration that the amount of experience which his subjects had with simple geometrical forms did not appreciably change the degree to which they spontaneously recognized these forms when they were embedded in complex figures. Apparently, in both the present investigation and in Gottschaldt's, the primary importance of the configurational influences is demonstrated.

2. *The Implications of the High Positive Correlations.*—It is highly significant that the performance of the subjects on any one of the main series of triangles correlated highly with their performance on either of the other two series of forms (See Table IV). As already stated, these correlations show the degree to which a subject's mean threshold tended to occupy on each of the three series of triangles the same relative position with regard to the distribution of individual thresholds of the various members of the group. The importance of these high positive correlations lies in the fact then that the subjects' general attitude (resulting from physiological condition, im-

mediately preceding experience, and the like) at the time of taking the test was not very important in determining the location of the closure threshold. If the subject's attitude, as just defined, had been of great importance in determining the location of his closure threshold, it would have resulted in variations from one test series to the next which, not being correlated with variations caused by similar factors in the case of the other subjects, would have greatly reduced the measured coefficients of correlation. The fact that these positive correlations are high indicates that the threshold for each series of triangles was dominantly determined by a relatively permanent organism-stimulus field relationship.

VII. Summary and Conclusions

A. Summary

The present study was designed as a first attack upon the experimental analysis of the phenomenon of closure. While closure effects have been noted incidentally in many previous investigations, the present work appears to be the first experimental study of closure *per se*. The general purpose of the study was that of determining whether or not, in the case of simple geometrical forms, the tendency toward closure could be psychophysically demonstrated to have a threshold characteristic, dependent upon a continuously variable aspect of the figure involved. Another purpose of the study was that of ascertaining whether or not the threshold of closure is the same for three different cases of the same geometrical form (*i.e.*, three different contour triangles). There were also more specific secondary objectives.

The threshold of closure was measured for three different kinds of triangles series, each ranging from very incomplete to very complete presentations. The threshold of closure was defined in terms of the percentage of the figure present at the point in the series marking the transition from forms having a twoness characteristic to those having a oneness characteristic. The factor of absolute size was controlled by use of a fourth triangle series.

B. Conclusions

1. The occurrence of closure, as defined and measured in this experiment, appears to follow a threshold function.

2. In the case of triangles and under the conditions of this experiment, the angular characteristics of the figure are important determinants of the location of the threshold. For the triangle series used here, the size of the threshold varied directly (though not necessarily proportionally) with the size of the indicated apex angle.

3. There is some indication that the absolute size of the figures involved affects the location of the threshold, as it was determined in this study; but the obtained difference was not in this case statistically significant. There is also the possibility that this difference is an expression of the fact that the larger forms bore a different relationship to the constant size exposure field than did the smaller figures.

4. Experience with partial representations of figures, within the limits of that given in the present experiment, does not significantly affect the threshold of closure. Rather, it is possible to consider the threshold as the expression of a dynamic stimulus field-organism relationship.

5. Since, under the conditions of the present study, the closure threshold is very stable, the subject's attitude of the moment appears to be relatively unimportant in determining the location of this point; and this value, under these conditions, must be construed, therefore, as an expression of a relatively permanent inter-relationship between the organism and the stimulus.

6. The explanation of the data of this experiment from an empirical point of view meets critical difficulties.

7. It is possible to frame an interpretation of the data of this study in terms of the principles of perception set forth by the Gestalt writers.

(Manuscript received August 14, 1941)

References

1. Collier, R. M., An experimental study of form perception in indirect vision, *J. comp. Psychol.*, 1931, **11**, 281–290.
2. Dodge, R., An improved exposure apparatus, *Psychol. Bull.*, 1907, **4**, 10–13.
3. Fuchs, W., Untersuchugen über das Sehen der Hemianopiker und Hemiamblyopiker. II: Die totalisierende Gestaltauffassung, *Zsch. f. Psychol.*, 1921, **86**, 1–143.
4. Gibson, J. J., The reproduction of visually perceived forms, *J. exp. Psychol.*, 1929, **12**, 1–39.
5. Gottschaldt, K., Über den Einfluss der Erfahrung auf die Wahrnehmung von Figuren. I. Über den Einfluss gehäufter Einprägung von Figuren auf ihre Sichtbarkeit in umfassenden Konfigurationen, *Psychol. Forsch.*, 1926, **8**, 261–317.
6. Gottschaldt, K., Über den Einflass der Erfahrung auf die Wahrnehmung von Figuren. II. Vergleichende Untersuchungen über die Wirkung figuraler Einprägung und den Einflus spezifisher Geschehensverläufe auf die Auffassung optischer Komplexe, *Psychol. Forsch.*, 1929, **12**, 1–87.
7. Hartmann, G. W., The concept and criteria of insight, *Psychol. Rev.*, 1931, **38**, 242–253.
8. Helson, H., The psychology of *Gestalt*, *Amer. J. Psychol.*, 1926, **37**, 25–62.
9. Hempstead, L., The perception of visual form, *Amer. J. Psychol.*, 1900, **12**, 185–192.
10. Koffka, K., *Perception:* an introduction to the *Gestalt-Theorie*, *Psychol. Bull.*, 1922, **19**, 531–581.
11. Koffka, K., *Principles of Gestalt psychology*, New York: Harcourt, Brace and Co., 1935.
12. Köhler, W., An aspect of Gestalt psychology, In Murchison, Carl (Ed.) *Psychologies of 1925*. Worchester, Mass: Clark Univ. Press., 1928, 163–195.
13. Liebmann, S., Über das Verhalten farbiger Formen bei Helligkeitsgleichheit von Figur und Grund, *Psychol. Forsch.*, 1927, **9**, 300–353.
14. Lindemann, E., Experimentelle Untersuchungen über das Entsehen und Vergehen von Gestalten, *Psychol. Forsch.*, 1922, **2**, 5–60.

15. MEAKIN, F., Mutual inhibition of memory images, *Psychol. Monog.*, 1903, **4**, 235–275.
16. OGDEN, R. M., Insight, *Amer. J. Psychol.*, 1932, **44**, 350–356.
17. OGDEN, R. M., *Psychology and education*, New York: Harcourt, Brace and Co., 1932.
18. REISER, O. L., The logic of Gestalt psychology, *Psychol. Rev.*, 1931, **38**, 359–368.
19. ROSENBLOOM, B. L., Configurational perception of tactual stimuli, *Amer. J. Psychol.*, 1929, **41**, 87–90.
20. STREET, R. F., A gestalt completion test. A study of a cross section of intellect, *Teach. Coll. Contrib. Educ.*, 1931, No. 481.
21. TIERNAN, J. J., The principle of closure in terms of recall and recognition, *Am. J. Psychol.*, 1938, **51**, 97–108.
22. WERTHEIMER, M., Untersuchungen zur Lehre von der Gestalt. II, *Psychol. Forsch.*, 1923, **4**, 301–350.
23. ZIGLER, M. J., An experimental study of visual form, *Amer. J. Psychol.*, 1920, **31**, 273–300.
24. ZIGLER, M. J., AND BARRETT, R., A further contribution to the tactual perception of form, *J. exp. Psychol.*, 1927, **10**, 184–195.
25. ZIGLER, M. J., AND NORTHUP, K. M., The tactual perception of form, *Amer. J. Psychol.*, 1926, **37**, 391–397.

THE RETENTION OF SERIAL LISTS OF ADJECTIVES OVER SHORT TIME-INTERVALS WITH VARYING RATES OF PRESENTATION [1]

BY ARTHUR W. MELTON AND G. RAYMOND STONE

University of Missouri and Brown University

Although reminiscence has attained the stature of an important phenomenon of serial rote learning as a consequence of the experimental studies of Ward (**26**) and Hovland (**10, 11, 12, 13, 14, 15**) and the elaborate theoretical interpretations of Hull *et al.* (**17**), the generality of the phenomenon in rote learning and the sufficient conditions for its production remain to be determined. The improvement of the recall of serial nonsense syllables after a rest-interval of two minutes, and a reduction in reaction latency, is a major prediction of Hull's theoretical system and a major result of the experimental studies. Whether the phenomenon reported in these experiments is, however, continuous with the improvements of retention reported by earlier investigators, and summarized by Woodworth (**27**) and McGeoch (**18**), cannot be determined. The Ward-Hovland reminiscence effect has been demonstrated only in the case of serial nonsense syllables which have been learned by massed practice with the anticipation method and with a two-second rate of syllable presentation, and the effect is maximal after a two-minute rest period. Earlier studies found reminiscence with both meaningful and meaningless materials, and the optimal intervals were frequently as long as two days. Many of these earlier studies, as Woodworth (**27**) remarks, did not eliminate either the possibility of rehearsal of the materials during the rest interval or the learning effect of the pre-rest test of performance. To date there have been no studies of reminiscence with meaningful materials with conditions and controls similar to those employed by Ward and Hovland.

The two experiments to be reported attempt to determine whether reminiscence occurs in the retention of serial lists of disconnected adjectives when the conditions of learning and retention, and the control of activity during rest intervals, are similar to those used by Ward and Hovland with nonsense syllables. The first experiment used the two-second rate of presentation previously used by Ward

[1] From the Laboratory of Psychology of the University of Missouri. We are indebted to the National Youth Administration for some of the subjects used in these experiments and for assistance in the tabulation of the data.

and Hovland in the successful demonstration of reminiscence with nonsense syllables. When reminiscence failed to occur, a second experiment was made in which the rate of presentation was increased to 1.45 seconds. Hovland (11) has shown that the faster rates of presentation favor the occurrence of reminiscence in the recall of nonsense syllables, and our two experiments with adjectives provide, therefore, a check on the generality of that specific conclusion.

Experimental Procedure

In Experiment I twenty-four college students learned lists of sixteen two-syllable adjectives until they correctly anticipated twelve or more adjectives. They then rested for the usual inter-trial interval of six seconds (continuous learning) or for two, five, or twenty minutes. They then relearned the list to a criterion of two successive errorless recitations. Throughout this experiment the rate of presentation of the adjectives was two seconds. In Experiment II another twenty-four college students learned the same lists under the same conditions, but the rate of presentation of the adjectives was 1.45 seconds throughout the learning and relearning. The normal inter-trial rest interval in this experiment was 5.9 seconds. In both experiments the control for the determination of the retention after the various intervals of rest was the condition in which the usual inter-trial interval elapsed between the attainment of the criterion of twelve correct anticipations and the relearning to the criterion of two successive errorless recitations.

The basic conditions and controls in the two experiments were identical. All subjects were given three days of practice in which they learned three lists of adjectives. On the first practice day all subjects learned one list to a criterion of two successive errorless recitations without rest; on the second day they learned a list to the same criterion with a two-minute rest after the trial on which twelve or more adjectives were correctly anticipated; and on the third day they learned a list to the same criterion with a two-minute rest between successive trials. This practice with distributed practice habituated the subjects to the shift from learning to color naming, which was the activity used to prevent rehearsal during the rest intervals.

After the three practice days all subjects served for four experimental days on which one list was learned and relearned under one of the four rest conditions. The four intervals of rest were counterbalanced for practice and inter-condition effects by a method of complete counterbalancing in which each of the twenty-four permutations of the four conditions served as the work-schedule for one subject. All subjects learned the same lists on the three practice and four experimental days. The variance due to the practice effect was, therefore, confounded with the variance due to lists, but every list was used equally often during the four experimental conditions.

The lists were made from two-syllable adjectives which were non-synonymous and non-antonymous. No list contained more than two adjectives with the same initial letter, and only three such duplications of initial letters were permitted in any list. No list contained more than four adjectives with the same terminal letter or syllable, and the frequencies of the various common adjectival endings were equalized as much as possible from list to list. Similar endings were not permitted in successive adjectives. Although these precautions against formal and meaningful similarity undoubtedly reduced the amount of intra-serial interference, they did not eliminate such interference altogether, if the presence of anticipatory and perseverative errors is taken as an indication of such interference.

The two-, five-, and twenty-minute rest intervals were filled with continuous color-naming. The subject was instructed to begin reading colors from a color-card as soon as the experimenter said 'colors,' and to lay aside the color-cards and prepare to anticipate the adjectives when the experimenter said 'stop.' The latter signal was given approximately six seconds before the cue for the first anticipation appeared in the aperture of the drum. The subject was urged to name the colors as fast as possible, and he was provided with twelve cards of 50 colors which were used in series. Ward (26) used a similar color-association activity in a sub-experiment to check on the control of the rehearsal factor in his major experiments, and found the reminiscence to be present after such color-association activity. Hovland (10) has regularly employed a somewhat less exacting color-naming activity in which the color squares are presented at the rate of one every two seconds.

All learning and relearning was by the anticipation method and in accordance with instructions which were similar to those regularly employed in experiments with nonsense syllables (21, 26). A difference exists between the instructions for anticipation in Ward's and Hovland's experiments. In the former the subject is instructed to anticipate the next unit in the list whenever possible, but he is not instructed to call out the unit when it appears in the aperture of the drum if he has been unable to anticipate it correctly. In Hovland's studies the latter instruction is given.[2] In the present studies Ward's type of instructions was used, because irrelevant interference of verbal responses may be introduced when the subject attempts to name the unit in the aperture of the drum and then anticipate the next unit in the list. During the practice days the subjects were given repeated instructions to attempt to anticipate as many adjectives as possible on the first relearning trial and to continue relearning with maximal effort until a criterion of two successive errorless recitations was achieved.

The adjectives were presented on a modified Hull drum driven by a synchronous motor. The latencies of all anticipations were determined by a DeSilva (4) microphone and electronic relay arrangement in which the relay was automatically reset when the drum turned. The recording polygraph operated at a speed of twenty inches per minute, which permitted measurement of latencies in units of .05 seconds. The latency measurements are known, from yet unpublished studies, to be sufficiently precise to reveal progressive decrements in latency as the amount of practice on serial lists of adjectives increases. They will be used in this study to determine whether increases or decreases in the latency of the association responses occur after the various intervals of rest.

Results

Rate of Presentation and the Difficulty of Learning.—The theorem employed by Hovland in formulating his studies of reminiscence (11) and distribution of practice (12) as a function of the rate of presentation of rote series was that as the magnitude of intra-serial inhibition decreases the amount of reminiscence and the effectiveness of distributed practice will decrease. Three lines of evidence for the greater intra-serial inhibition with the faster rate of presentation were: (1) the greater number of trials required for mastery to a criterion of partial (11) or complete (12) learning, (2) the greater increase in the number of errors on the middle units than on the end units, and (3) the greater range of 'oscillations' at the threshold of recall, as measured by the difference between the number of presentations of each unit (a) preceding the first correct anticipation, and (b) including the last failure to anticipate. The more elaborate theoretical analysis of Hull *et al.* (17) uses these three measures as standard indices of intra-serial inhibition, and it is important, therefore, to see whether change of the rate of presentation of the adjetives in these experiments was accompanied by a change in the amount of intra-serial inhibition as measured by them.[3]

[2] Hovland followed Hull (16) in this variation from the usual instructions for the anticipation method. It should be noted that the subjects called out the syllable in the aperture both when they had anticipated it incorrectly and when they had made no attempt to anticipate it (personal communication from Dr. Hovland). The significance of this methodological variation is not clear. Both Ward (26) and Hovland (10, 11) obtained reminiscence, but there were other differences between their experiments, *e.g.*, different lists of syllables, spelling the syllables *versus* pronouncing them, and the activities used to fill the rest intervals.

[3] Hovland does not consider the use of total learning time (number of units × presentation time × trials) as a measure of difficulty. He restricts his discussion to trials or presentations. It is well known from the work of Gamble (7, 8) that the faster rates of presentation, at least

The frequently noted increase in the number of trials for mastery when the rate of presentation is increased is again verified in this experiment. The subjects in Experiment I learned the sixteen adjectives to a criterion of two successive errorless recitations in 20.21 ± 2.36 trials, with the two-second rate; and the subjects in Experiment II learned in 23.12 ± 1.87 trials, with the 1.45-second rate.[4] Similar results were obtained when the learning records up to the criterion of twelve correct anticipations were pooled for all conditions in the two experiments. In this case the mean trials for learning with the two-second rate was 9.48 ± .92, and the mean trials for learning with the 1.45-second rate was 10.40 ± .62. These differences are obviously unreliable, but it will be noted in Table I that the results for the several conditions of each experiment are consistent with the pooled results.

The other two measures of rate of learning which are assumed to be indices of intra-serial inhibition are summarized in Fig. 1. The mean number of recall failures during continuous learning to two successive errorless trials is greater for the middle units of the lists with the 1.45-second rate than with the two-second rate, but no such difference occurs for units 1 through 5 and 14 through 16. The same trend is present when the data of the four conditions at each rate are combined to determine the mean frequency of recall failures up to the criterion of twelve correct anticipations. None of these differences is statistically significant. Thus, the mean difference

within the range from four seconds to two seconds, increase the number of repetitions for learning, but decrease the total learning time. Hovland's data are not conclusive on this point, but they indicate that repetitions for learning may vary greatly while total learning time varies but little. Thus, in his study of reminiscence (11) the trials to a criterion of seven correct anticipations were 6.05 with the two-second rate and 3.28 with the four-second rate, which means that the mean time spent in learning was approximately 145.2 seconds with the former rate and 154.4 seconds with the latter rate. In his study of distribution of practice (12) the trials to learn to one errorless recitation with the two-second and four-second rates were 14.89 and 6.78, respectively. The mean times spent in learning were approximately 357.4 seconds and 325.4 seconds. There is some evidence in his data (12, Table 4) that the faster rate of presentation increases the trials to learn to all criteria of partial mastery of the list, but that it increases the amount of learning time only for criteria of mastery beyond nine correct anticipations in the twelve-unit list. It should also be noted that the supposed greater difficulty of the middle units of the list with the faster rate of presentation (11, Fig. 1) disappears altogether when learning time, instead of repetitions, is used to measure difficulty. On the other hand, the difference between massed and distributed practice is greater for the faster rate of presentation, whichever measure of difficulty is used. Hovland's observed correlations between rate of presentation and reminiscence and the effectiveness of distributed practice are not, of course, called in question. One must, however, question the practice of using one measure of learning difficulty as an index of intra-serial inhibition when another equally valid method of measurement yields contradictory evidence of difficulty. Fortunately, with the precise control of presentation times in current studies of rote learning, the data on trials or presentations can be readily converted into learning times.

[4] All measures of reliability are sigmas. All sigmas of the mean differences have been corrected for the correlation between the arrays when such a correction was appropriate. For this experimental design there are twenty-three degrees of freedom and a *t*-value of 2.807 is accepted as an indication of a very significant difference. Cf. (6).

between the recall failures on items 6 through 13 was only 0.84 ± .88 per unit (t = 0.96) for learning up to the criterion of twelve correct anticipations, and only 1.46 ± 1.76 per unit (t = 0.83) for learning to complete mastery. The agreement between the two sets of curves suggests, however, the validity of the difference.

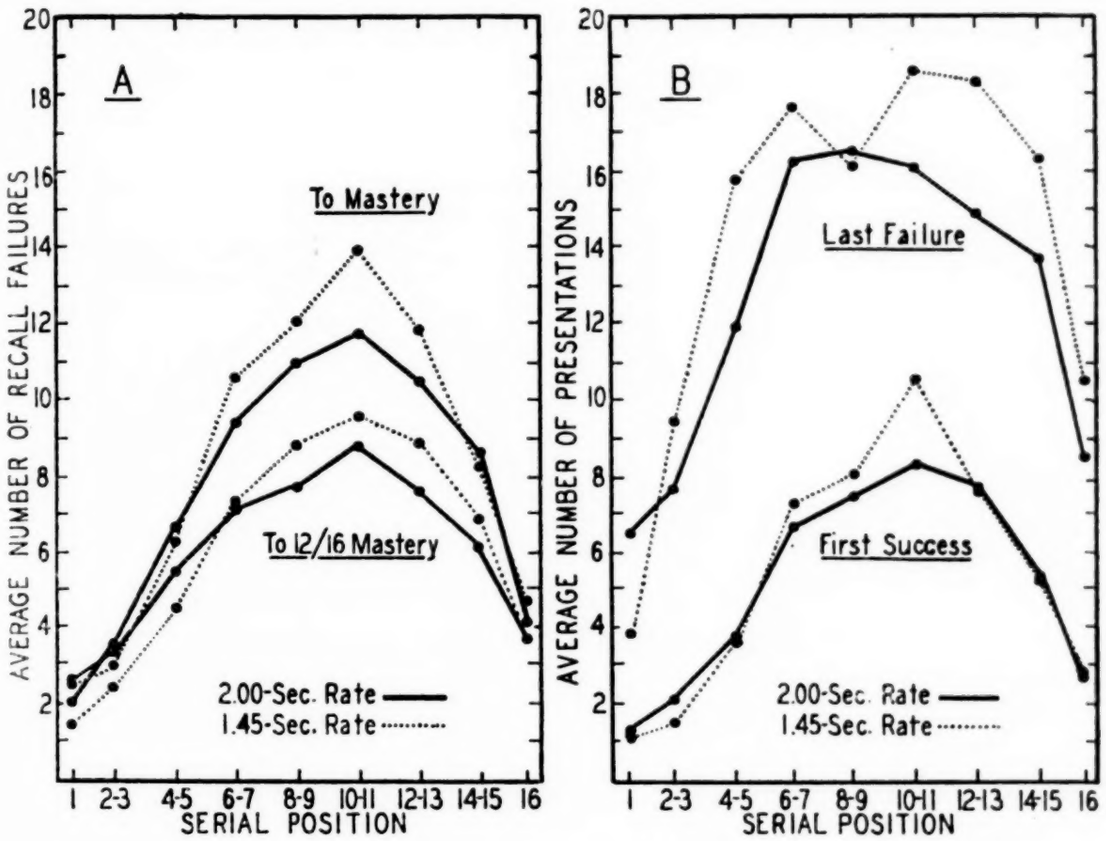


FIG. 1. The difficulty of learning as a function of serial position and rate of presentation, as measured by (*A*) the number of recall failures up to 12/16 mastery and to two successive errorless trials, and (*B*) the number of presentations before the first successful recall and before the last failure to recall correctly.

The results with the measures which delimit the period of 'oscillation' of recall for each serial unit are also in agreement with Hovland's findings with nonsense syllables. As shown in section B of Fig. 1, the greater difficulty of the middle units in the lists with the 1.45-second rate is more a difficulty of permanent fixation, as represented in the upper curves, than it is a difficulty of making the first correct anticipation, as represented in the lower curves. The measures were obtained from records of twenty-four subjects during continuous learning to two successive errorless trials.

The reaction times obtained during the present experiment provide still another measure of the effects of change in rate of presentation, if reaction latency may be assumed to be positively correlated with the inhibition to which a serial unit is subjected (17). In Fig. 2

are the mean latencies of correct anticipations in the various serial positions on the first trial during which twelve or more adjectives were correctly anticipated, and on the first of two successive errorless trials when learning was continuous. These data are ambiguous. At the attainment of the criterion of partial learning, the two-second rate seems to give a much more pronounced increase in the latency of the middle units than the 1.45-second rate, which contradicts the expectation. Furthermore, mean latency of all correct anticipations at the two-second rate is .807 ± .027 sec. and the mean latency of

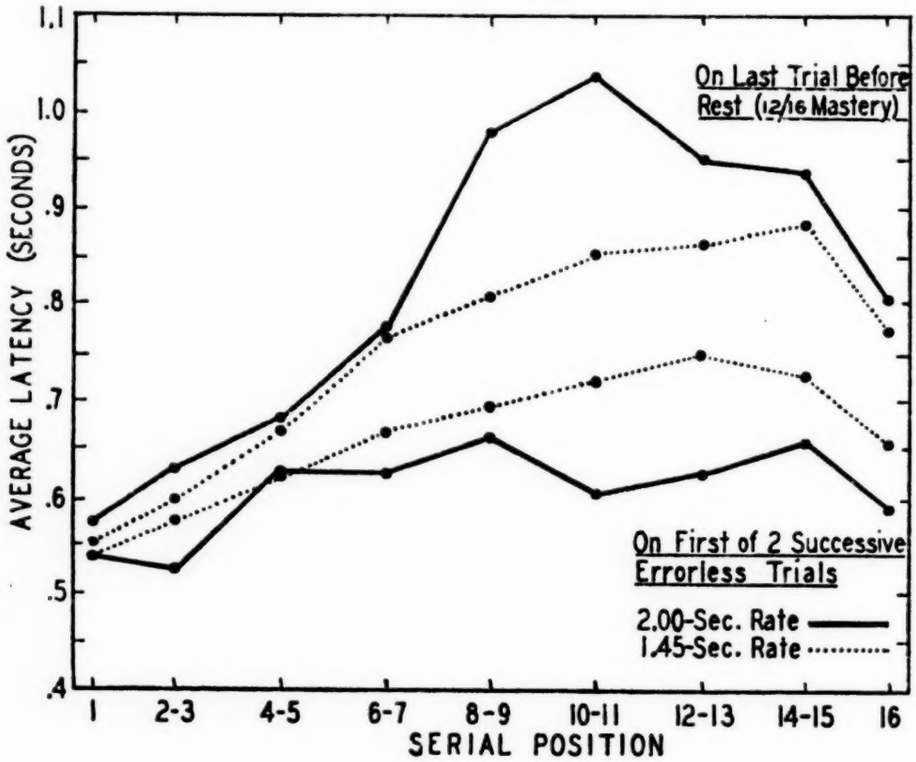


FIG. 2. The average latency of correct anticipations on the last trial before the rest period (12/16 mastery) and on the first of two successive errorless trials, as a function of the serial position and rate of presentation of the adjectives.

the correct anticipations at the 1.45-second rate is .739 ± .023 sec. (Diff. = .068 ± .035; t = 1.94). On the first of two successive errorless trials, these relations are apparently reversed, largely as a consequence of the virtual disappearance of latency variations from serial position to serial position in the case of the two-second rate. The mean latencies at the two-second and 1.45-second rates are .611 ± .037 and .709 ± .036 (Diff. = .098 ± .052; t = 1.89). Our limited knowledge of the factors which determine reaction latency in serial learning makes interpretation of these differences impossible. It is sufficient for the present to know that the other three indices of intra-serial inhibition employed by Hovland indicate an increase

in such inhibition when the rate of presentation was increased from two seconds to 1.45 seconds.

Learning Before the Various Rest Intervals.—In Table I are shown (*a*) the mean trials required to attain the criterion of twelve correct anticipations under each condition of the two experiments, (*b*) the mean correct anticipations on the criterial trial, and (*c*) the mean latency of correct anticipations on the criterial trial. These data warrant the conclusion that there was no significant variation of the

TABLE I

DEGREE OF LEARNING OF LISTS BEFORE INTRODUCTION OF REST PERIODS

Condition	Mean Trials to Criterion of 12 Correct Units		Mean Correct Units on Criterial Trial		Mean Latency (Sec.) of Correct Units	
	2.00-sec. Rate	1.45-sec. Rate	2.00-sec. Rate	1.45-sec. Rate	2.00-sec. Rate	1.45-sec. Rate
6-sec. rest	10.29±1.54	10.63±0.75	12.63±.13	12.63±.15	.81±.04	.74±.03
2-min. rest	9.50±1.06	10.33±0.81	12.54±.15	12.54±.18	.81±.03	.76±.04
5-min. rest	8.58±0.74	10.50±1.01	12.67±.13	12.54±.14	.82±.03	.73±.03
20-min. rest	9.54±0.92	10.13±0.73	12.79±.18	12.79±.20	.76±.03	.73±.03

degree of learning of the lists before the introduction of the rest periods of various durations. With the two-second rate of presentation, the largest difference between the mean trials to learn was 1.71 ± 1.31 (t = 1.31), and with the 1.45-second rate the largest difference was .50 ± .71 (t = 0.70). The largest difference between the mean correct anticipations on the criterial trial was .25 ± .20 (t = 1.25) in both experiments. The largest difference between the mean reaction latencies was .06 ± .034 sec. (t = 1.76) with the two-second rate and .03 ± .029 sec. (t = 1.03) with the 1.45 second rate.

Effects of Rest Periods on Recall.—Although the phenomenon of reminiscence and the effectiveness of the distribution of practice are closely related (**2, 3**) reminiscence is properly defined as an improvement in performance immediately after an interval of no practice (recall); whereas, the effectiveness of distributed practice is revealed primarily by the trials or time or errors for mastery. This distinction is justified by the history of experimental work on reminiscence and on the effectiveness of distributed practice, as well as by the discovery that under some circumstances the effects of a rest period on recall and relearning vary independently (**23**). Our data will be presented in accordance with this distinction.

The primary data for the determination of the reminiscence effect are presented in Table II. The recall scores, that is, the number of correct anticipations on the first relearning trial, fail to show improvement after any interval of rest. Instead, with the two-second rate

of presentation the mean recall after two and five minutes of rest was only 92 percent of the recall after six seconds, and the recall after twenty minutes was only 69 percent. Furthermore, no absolute reminiscence occurred after any interval of rest when the rate of presentation was increased to 1.45 seconds. In this case the retention after two, five, and twenty minutes was 99.3, 95.7, and 66.8 percent. The *t*-values of the differences between the six-second rest condition and the two-, five-, and twenty-minute rest conditions in

TABLE II

AMOUNT OF RETENTION AFTER VARIOUS INTERVALS OF TIME AS MEASURED BY RECALL SCORES

Condition	Mean Correct Anticipations on the First Relearning Trial		Mean Correct Anticipations on the Second Relearning Trial	
	2.00-sec. Rate	1.45-sec. Rate	2.00-sec. Rate	1.45-sec. Rate
6-sec. rest	12.42±.46 100.0%	11.54±.42 100.0%	12.75±.51 100.0%	12.63±.37 100.0%
2-min. rest	11.46±.40 92.3%	11.46±.45 99.3%	12.58±.45 98.7%	12.25±.49 97.0%
5-min. rest	11.46±.50 92.3%	11.04±.38 95.7%	12.46±.41 97.7%	11.08±.56 87.6%
20-min. rest	8.63±.60 69.6%	7.71±.68 66.8%	11.04±.55 86.6%	9.29±.68 73.4%

Experiment I are 1.57, 1.65, and 5.83, respectively. The corresponding *t*-values in Experiment II are 0.13, 1.08, and 4.45. The forgetting with the shorter intervals is not statistically significant; but it is forgetting and not reminiscence. It will also be noted that the mean correct anticipations on the second relearning trial confirm in every instance the presence of forgetting.

A count of the number of reminiscent items in the records of all subjects on the first recall trial likewise failed to reveal a significant increase in their number with the two-minute and five-minute rest intervals. A reminiscent item is defined as an item correctly anticipated on the first relearning trial without having been correctly anticipated on any learning trial. The number of such items was 12, 19, 12, and 6 after the six-second, two-minute, five-minute, and twenty-minute rest intervals with the two-second rate of presentation. The frequencies were 16, 7, 6, and 1 after the corresponding rest intervals with the 1.45-second rate of presentation. The apparent increase in the frequency of reminiscent items after the two-minute rest with the two-second rate is obviously unreliable, and a very large number of subjects would be required in order to give the data an opportunity

to disprove the hypothesis that the rest interval has no effect on the frequency of such items. In addition, the experiment with the faster rate of presentation, which should have increased the reminiscent items, gives no evidence for such an increase.

Two additional indices of recall performance further demonstrate this forgetting. Hull *et al.* (**17**) have deduced from a special theory of reminiscence that the effect, if present, will occur maximally in the recall of the middle items of a rote serial list. Verification of this

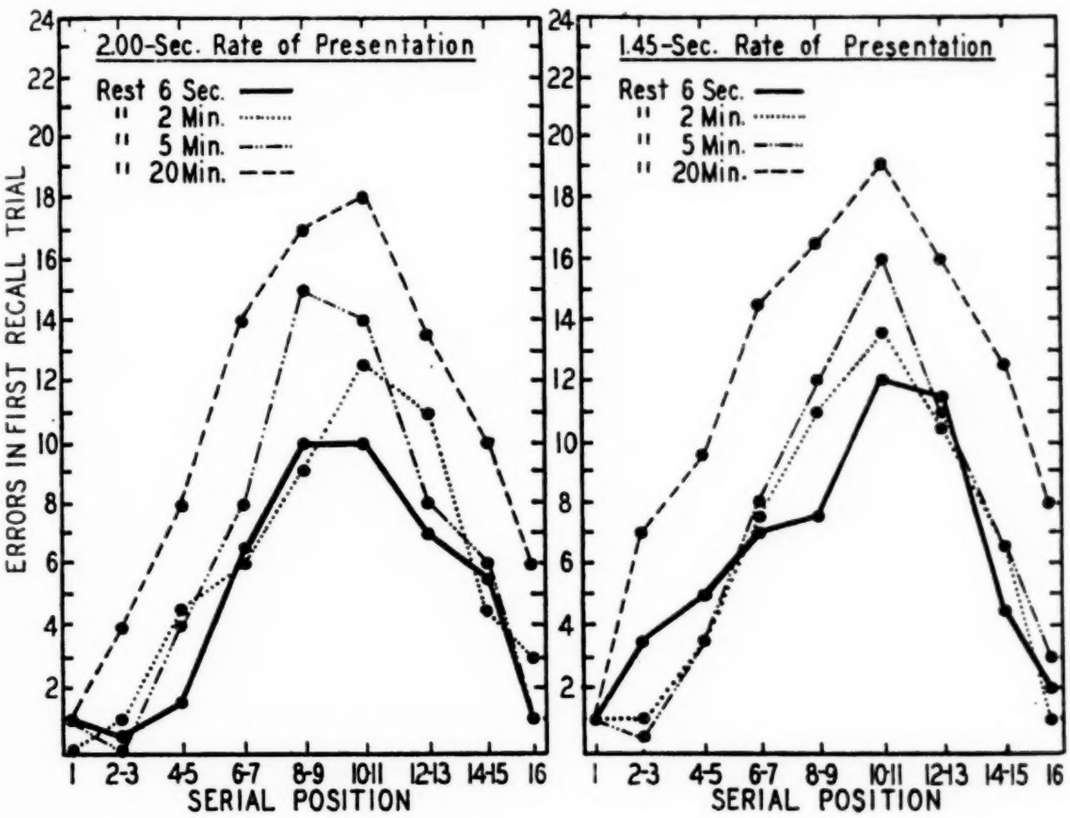


FIG. 3. The number of failures to recall adjectives in the various serial positions on the first recall (relearning) trial, as a function of the duration of the rest interval after 12/16 mastery and the rate of presentation of the adjectives.

has been reported by Hovland (**10**). In Fig. 3 are shown the frequencies of errors in recall on the first relearning trial at the various serial positions when the rates of presentation were two seconds and 1.45 seconds. If a relative reminiscence effect were present, the curves for the two-minute and/or five-minute rest conditions should fall below the solid-line curve for the continuous learning condition in the middle range of the list, *e.g.*, from units 6 to 13. It is clear that such relative reminiscence of the middle items failed to occur in either experiment. On the contrary, the greatest absolute forgetting occurred among the middle items after all rest intervals.[5]

[5] We have refrained from computing ratios of recall of middle and end items (**10, 16**). The meaning of such ratios is ambiguous, at least in any experiment where the subjects are carried

The other index of recall performance accepted by Hull *et al.* (17) as an operation for the measurement of reminiscence is the latency of the correct anticipations on the first relearning trial. The assumed inverse relationship between effective excitatory strength and latency of response is founded on Simley's (25) work with paired associates, but the present results (Fig. 2) suggest that the assumption is valid for serial learning as well. Ward (26) presents evidence for a decrease in reaction latency with rest intervals which are optimal for reminiscence in the recall of nonsense syllables. In Fig. 4 the

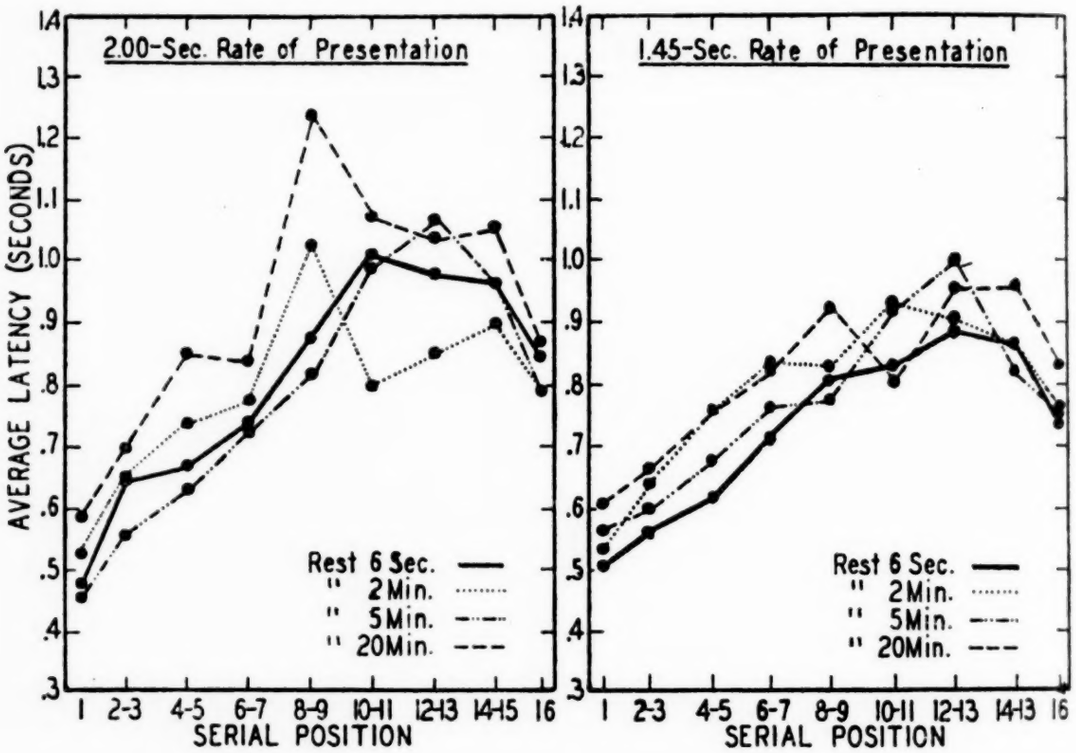


FIG. 4. The average latency of correct recalls of adjectives in the various serial positions on the first recall (relearning) trial, as a function of the duration of the rest interval after 12/16 mastery and the rate of presentation of the adjectives.

mean latency of all correct responses of all subjects at a particular serial position is plotted against the serial position of the unit recalled. Again the data for the continuous learning condition are represented by the solid line. With the two-second rate of presentation there is an apparent reduction of the latency of correct responses in positions 10 through 16 when the rest period was two minutes. This is, however, the only evidence in either experiment which seems to indicate a reminiscence effect, and its significance may be questioned. In the

beyond the mastery of 50 percent of the material before the rest interval is introduced. Improvement of a significant amount in the recall of items near the ends of the list is impossible. If, therefore, a reminiscence effect is found in the total recall score, that effect could occur only in the recall of the middle items of the list.

first place, the irregularity of the latency curve for the two-minute rest condition in Experiment I suggests a low reliability of latency measurements, especially for middle units of the list.[6] Secondly, the deviations from a regular curve form are much less pronounced in Experiment II, and in this experiment the latencies are regularly greater after the two-minute rest than after the six-second rest, even though the two-minute rest in this experiment gave the nearest approach to reminiscence in terms of absolute recall scores.

TABLE III
AMOUNT OF RETENTION AFTER VARIOUS TIME-INTERVALS AS MEASURED BY RELEARNING SCORES

Condition	Mean Relearning Trials to 1 Perfect Recitation		Mean Relearning Trials to 2 Successive Perfect Recitations	
	2.00-sec. Rate	1.45-sec. Rate	2.00-sec. Rate	1.45-sec. Rate
6-sec. rest	6.67±1.11 100.0%	7.42± .89 100.0%	8.92±1.18 100.0%	11.50±1.34 100.0%
2-min. rest	6.29± .91 94.3%	7.00±1.06 94.3%	9.42±1.18 105.2%	11.96±1.30 104.0%
5-min. rest	5.54± .59 83.1%	8.25±1.42 111.2%	9.04±1.11 101.3%	14.58±1.77 126.8%
20-min. rest	7.71± .75 115.6%	12.29±1.55 165.6%	10.79±1.25 121.0%	18.71±2.13 162.7%

Effects of Rest Periods on Relearning to Mastery.—In Table III are presented the mean trials required for relearning to one perfect recitation and to two successive perfect recitations after the various intervals of rest. Although relearning to the less stringent criterion is apparently more rapid with the two- and five-minute rest intervals than with the six-second rest interval (continuous learning) when the rate of presentation is two seconds, neither of these decreases in trials to relearn is reliable. With the five-minute rest the difference is 1.13 ± 1.18 trials (t = 0.96). With the 1.45-second rate of presentation the two-minute rest is the only rest interval which led to fewer relearning trials to one perfect recitation, and the decrease in this case is obviously unreliable (t = 0.39). All conditions of extended rest required more relearning trials to the criterion of two successive perfect recitations than the condition in which the usual inter-trial rest of six seconds was given.

[6] We know of no accurate method for evaluating the reliability of mean differences in latency when the same subjects do not yield latency measurements for a particular item under all experimental conditions, or when the number of latency measurements for a particular trial varies from subject to subject and from condition to condition. This is a severe limitation on the usefulness of latency measurements as a sensitive indicator of retention or inhibition.

In general, it may be said that the effect of a rest period introduced after three-fourths of the material has been learned is beneficial, if at all, only when the criterion of mastery is one perfect recitation, and it is more likely to be beneficial if the rate of presentation is two seconds than it is if the rate of presentation is 1.45 seconds. The latter conclusion is clearly at variance with the expectation based on the theory of Hull *et al.* (**17**), in view of our earlier evidence for a greater accumulation of inhibition during the original learning at the 1.45-second rate.

Discussion

It must be concluded that reminiscence failed to occur in the recall of partially learned serial adjectives after two, five, and twenty minutes of no practice. In addition, it has been shown that an increase in the rate of presentation from two seconds to 1.45 seconds per adjective did not produce reminiscence or reduce significantly the amount of forgetting.

These results further limit the range of conditions under which reminiscence has been obtained in rote learning. Hovland has shown that reminiscence is not present when serial nonsense syllables are presented at a four-second rate (**11**), when serial nonsense syllables are learned at a two-second rate but with two minutes between trials (distributed practice) (**10**), and when paired nonsense syllables are presented at the rate of four seconds per pair and the serial order of the pairs varies from trial to trial (**13**). Also, Shipley (**24**) obtained no reminiscence when lists of eight, fourteen, and twenty serial nonsense syllables were presented at a 2.3-second rate with massed practice, even though Hovland's other conditions were duplicated. To date, therefore, reminiscence after short intervals of no practice has been obtained only in the studies of Ward (**26**) and Hovland (**10, 11, 14**) in which serial lists of twelve nonsense syllables have been learned at a two-second rate with six seconds between trials. Although these positive and negative results may eventually be integrated by a systematic theory such as proposed by Hull *et al.* (**17**), a defensible position at this time is that the essential conditions of the phenomenon are unknown and further systematic exploration should precede a formal theory.

This noncommittal position is further justified when the present results are considered in the light of the basic tenet of the response-interference theory which seems most adequate for the interpretation of the results of Ward and Hovland. This basic proposition is that reminiscence appears when there is a significant amount of intra-habit response interference at the termination of the original learning. Intra-serial interference is, however, a prominent feature of the rote

learning of serial adjectives, as indirectly revealed by the greater difficulty of the middle units in the lists and by McGeoch's (**19**) evidence that remote associations, which are the supposed basis for some, at least, of the intra-serial interference, occur with high frequency in the serial learning of adjectives. The relatively greater increase in the difficulty of the middle units in the adjective lists following the increase in the rate of presentation likewise indicates that intra-serial interference was present and subject to variation in amount, but neither rate of presentation resulted in reminiscence. Under these circumstances it serves no purpose to conclude that the differentiating factor in these experiments and the experiments of Ward and Hovland is the meaningfulness of the material learned, especially in view of the fact that the present experiments duplicate in detail only the methods and controls employed by Ward in a supplementary experiment with six subjects (**26**, p. 41).

Although our conclusions are negative when viewed with reference to the phenomenon of reminiscence, there are positive implications of some importance for studies of retention and forgetting. In the first place, the rate of forgetting of the adjectives during the twenty minutes following practice is accelerating, rather than decelerating, as revealed by the fact (Table II) that the decrement in recall after twenty minutes is more than four times the decrement after five minutes. The data do not, therefore, fit an extrapolated Ebbinghaus (**5**) curve. This acceleration of forgetting is likewise reported by Hovland (**15**) for serial nonsense syllables learned by distributed practice, even though no reminiscence occurred. It is probable that the other conditions which have yielded no reminiscence also produced an accelerating forgetting curve, judging from the very small amounts of forgetting recorded after the two-minute rest periods used.

This acceleration of forgetting during the first few minutes after practice may require an interpretation continuous with an interpretation of reminiscence, but it may, on the other hand, be the result of factors which could never produce an *increase* in recall. Thus, the color-naming activity used to fill the 'rest' interval may produce ennui or a disturbance of the set for anticipatory recall in amounts which increased at an accelerating rate, after the manner of a work-decrement, as the color-naming continued. Such a circumstance could account for an acceleration of forgetting, but could never account for an absolute reminiscence effect.

A second positive implication of some importance is that the many investigations of the retention of rote meaningful materials over short intervals of time probably have not been complicated by the phenomenon of reminiscence. This is especially pertinent to the many studies of retroactive inhibition by McGeoch (*e.g.*, **20**) and others

(1) in which lists of adjectives have been learned under essentially the same conditions used in these experiments. Were reminiscence present in the recall of adjectives, the conditions for reminiscence of the first set of responses during the learning of the second set would be near-optimal in such studies, since the interval of rest between the learning of the original and interpolated lists is usually between thirty seconds and two minutes. The complication of the usual inter-list interference effects by an increase in the effective strength of the first set of responses would influence significantly the role of both generalization (9) and unlearning (21, 22) in retroactive inhibition. Since both adjectives and nonsense syllables are commonly employed in studies of retroactive inhibition, it follows that the interpretation of studies of retroactive inhibition will in an important sense depend upon the results of further systematic exploratory studies of short-time retention.

Summary

1. Twenty-four subjects learned serial lists of sixteen adjectives to a criterion of twelve correct anticipations, and then rested for the usual inter-trial interval of six seconds, or for two, five, or twenty minutes, after which they relearned the lists to a criterion of two successive errorless trials. The rate of presentation was two seconds per adjective. In a second experiment another twenty-four subjects learned under the same conditions except for the rate of presentation, which was increased to 1.45 seconds per adjective. The primary purpose was to discover whether reminiscence occurred in recall after the intervals of rest longer than six seconds, and whether the amount of reminiscence was a function of the rate of presentation of the adjectives.

2. An increase in the rate of presentation of the adjectives (*a*) increased the trials required for partial and for complete mastery of the lists, but did not increase the total time spent in learning; (*b*) increased the number of presentations for mastery of the middle items in the lists more than for the end items; (*c*) increased the 'oscillations' of recall; and (*d*) affected the latency of the correct anticipations. These effects supposedly indicate that the increase in the rate of presentation increased the amount of intra-serial inhibition during learning.

3. Some decrement in recall, rather than reminiscence, was found after all extended intervals of rest with both rates of presentation. There was, furthermore, no evidence of a relative reminiscence effect, since (*a*) the latencies of correct anticipations after the extended intervals of rest were, on the whole, longer than the latencies on the corresponding trial during continuous learning, and (*b*) after the ex-

tended intervals of rest the middle items of the lists did not show a relatively smaller decrement in recall than the end items.

4. Although reminiscence was not obtained, the amount of forgetting, as measured by recall, during the first twenty minutes after the cessation of practice increased at an accelerating rate.

5. The effect of a rest period on the trials required for relearning to mastery is beneficial, if at all, only when the criterion of mastery is one errorless trial, and it is more likely to be beneficial if the rate of presentation is two seconds than it is if the rate is 1.45 seconds. With these restrictions, the two- and five-minute rest intervals are optimal.

6. These results further limit the range of conditions under which reminiscence occurs in rote learning. It is concluded that further systematic exploration of the conditions which affect the retention of rote verbal materials is required before a formal integrating theory is attempted.

7. These results have significance for the studies of retroactive inhibition which have used similar materials, since the presence or absence of reminiscence should influence the operation of factors responsible for retroactive inhibition.

(Manuscript received August 5, 1941)

References

1. Britt, S. H., Retroactive inhibition: A review of the literature, *Psychol. Bull.*, 1935, **32**, 381–440.
2. Bunch, M. E., The measurement of reminiscence, *Psychol. Rev.*, 1938, **45**, 525–531.
3. Bunch, M. E., & Magdsick, W. K., The retention in rats of an incompletely learned maze solution for short intervals of time, *J. comp. Psychol.*, 1933, **16**, 385–409.
4. DeSilva, H. R., A timer for vocal reactions with thyratron tubes and microphones, *Amer. J. Psychol.*, 1935, **47**, 315–317.
5. Ebbinghaus, H., *Memory* (Tr. by H. A. Ruger and C. E. Bussenius), New York: Teachers College, 1913.
6. Fisher, R. A., *Statistical methods for research workers*, London: Oliver and Boyd, 1934, (5th ed.), p. 158.
7. Gamble, E. A. McC., Two studies in memorizing by slow and by rapid repetition. I. Rate of presentation and tenacity of impression, *Psychol. Monogr.*, 1916, **22**, 98–151.
8. Gamble, E. A. McC., A study of three variables in memorizing, *Amer. J. Psychol.*, 1927, **39**, 223–234.
9. Gibson, E. J., Retroactive inhibition as a function of degree of generalization between tasks, *J. exp. Psychol.*, 1941, **28**, 93–115.
10. Hovland, C. I., Experimental studies in rote-learning theory. I. Reminiscence following learning by massed and by distributed practice, *J. exp. Psychol.*, 1938, **22**, 201–224.
11. Hovland, C. I., Experimental studies of rote-learning theory. II. Reminiscence with varying speeds of syllable presentation, *J. exp. Psychol.*, 1938, **22**, 338–353.
12. Hovland. C. I., Experimental studies of rote-learning theory. III. Distribution of practice with varying speeds of syllable presentation, *J. exp. Psychol.*, 1938, **23**, 172–190.
13. Hovland, C. I., Experimental studies of rote-learning theory. IV. Comparison of reminiscence in serial and paired-associate learning, *J. exp. Psychol.*, 1939, **24**, 466–484.
14. Hovland, C. I., Experimental studies in rote-learning theory. V. Comparison of distribution of practice in serial and paired-associate learning, *J. exp. Psychol.*, 1939, **25**, 622–633.

15. HOVLAND, C. I., Experimental studies in rote-learning theory. VI. Comparisons of retention following learning to same criterion by massed and distributed practice, *J. exp. Psychol.*, 1940, **26**, 568–587.
16. HULL, C. L., The influence of caffeine and other factors on certain phenomena of rote learning, *J. gen. Psychol.*, 1935, **13**, 249–274.
17. HULL, C. L., HOVLAND, C. I., ROSS, R. T., HALL, M., PERKINS, D. T., & FITCH, F. M., *Mathematico-deductive theory of rote learning—a study in scientific methodology*, New Haven: Yale Univ. Press, 1940.
18. MCGEOCH, G. O., The conditions of reminiscence, *Amer. J. Psychol.*, 1935, **47**, 65–89.
19. MCGEOCH, J. A., The direction and extent of intra-serial associations at recall, *Amer. J. Psychol.*, 1936, **48**, 221–245.
20. MCGEOCH, J. A., Studies in retroactive inhibition: VII. Retroactive inhibition as a function of the length and frequency of interpolated lists, *J. exp. Psychol.*, 1936, **19**, 674–693.
21. MELTON, A. W. & IRWIN, J. MCQ., The influence of degree of interpolated learning on retroactive inhibition and the overt transfer of specific responses, *Amer. J. Psychol.*, 1940, **53**, 173–203.
22. MELTON, A. W. & VON LACKUM, W. J., Retroactive and proactive inhibition in retention: Evidence for a two-factor theory of retroactive inhibition, *Amer. J. Psychol.*, 1941, **54**, 157–173.
23. MELTON, A. W., Differential effects of rest pauses on reminiscence and relearning in the pursuitmeter habit (in preparation).
24. SHIPLEY, W. C., The effect of a short rest pause on retention in rote series of different lengths, *J. gen. Psychol.*, 1939, **21**, 99–117.
25. SIMLEY, O. A., The relation of subliminal to supraliminal learning, *Arch. of Psychol.*, 1933, **22**, (No. 146), pp. 40.
26. WARD, L. B., Reminiscence and rote learning, *Psychol. Monogr.*, 1937, **49**, (No. 220), pp. 64.
27. WOODWORTH, R. S., *Experimental psychology*, New York: Henry Holt & Co., 1938.

A STUDY OF THE FITNESS OF COLOR COMBINATIONS, IN DUPLE AND IN TRIPLE RHYTHM, TO LINE DESIGNS

BY IVY G. CAMPBELL

Wells College

The main problem investigated in the following experiments was: "Do Os agree in their judgments as to the fitness of certain color combinations to certain designs?" A subsidiary problem was: "Does rhythm play a role in these judgments?" On the basis of experiments that have been performed, notably by Koffka [1] and by Ruckmick [2] we assumed that visual rhythm is a fact. And we raised the problem whether certain types of line designs in certain color combinations are more fitting to certain types of rhythm than are others. From preliminary experiments we received strong suggestion that this is true and in order to test this as well as to secure data on our main problem the following experiment was designed and performed in collaboration with Miss Jean Lewis, Wells 1932.

Three designs having the dominant characteristics of rectangular, curving, sharp-angled,[3] respectively, were chosen. As our problem was to determine the effect both of type of line design and of rhythm, designs which would clearly present rhythm had to be found. Rhythm we held to be present when there were repeating groups made up of identical members but within each group accent given to one member by some means of emphasis. Mere repetition of members did not, we held, give rhythm. We therefore chose designs with the same shape motive regularly repeated in a continuous horizontal series and put rhythm—duple and triple—into them by coloring the motives in such a way as to give repeating identical groups. Each group contained 2 colors; for the duple rhythm these 2 colors merely alternated, for the triple rhythm a motive was drawn in 1 color followed by 2 repetitions of the motive in another color. The following color-combinations were used: Bl-R (black and red), B-G (blue and green), B-Y (blue and yellow), G-R (green and red), Bl-Y (black and yellow), G-Y (green and yellow).

In Part I the 3 designs, motive repeated 6 times (plate I), were drawn on cards of white paper, 6 in. × 4 in. They were colored with Eagle Colored School Pencils in duple rhythm in each of the 6 color-combinations. Equated with the Munsell Color Book the colors were: Black 3; Red 5/10; Blue = B-PB 7/6; Green = GY-G 7/4; Yellow 8/10.

The procedure of Part I was as follows: Os were admitted 1 at a time into a room where on a couch covered with a white cloth the 18 designs were laid out in the following order:

	Black and Red	Blue and Green	Blue and Yellow	Green and Red	Black and Yellow	Green and Yellow
Rectangles	"	"	"	"	"	"
Curves	"	"	"	"	"	"
Angles	"	"	"	"	"	"

[1] Koffka, K., Experimental-Untersuchungen zur Lehre von Rhythmus, *Z. Psychol.*, 1909, 52, 1–109.

[2] Ruckmick, C. A., Visual rhythm, in *Studies in psychology:* Titchener commemorative volume, Louis N. Wilson, Worcester, Mass., 1917, 231–254.

[3] In this article these will be referred to as angle designs or angles.

The Os sat facing the couch and gave their judgments orally to the experimenter. The instructions given orally to the Os were: "You will be shown a series of designs in different color-combinations. Disregard, as much as possible, your liking for both the designs and the colors. In the first group, the black and red, choose the design which seems to you to be the most fitting to the color-combination. Do likewise for each color-combination group." This choice will be referred to as Design Choice. When these judgments had been made, the following directions were read for the second part of the experiment: "Now that the judgments on the color groups have been made, I would like a similar judgment on the design groups. This time choose the color-combination which seems to you to be most fitting to the rectangular design, then the curve design, then the angle design." This choice will be referred to as Color Choice. The observations were always made on a sunny afternoon. Altogether 62 Os made judgments but these 62

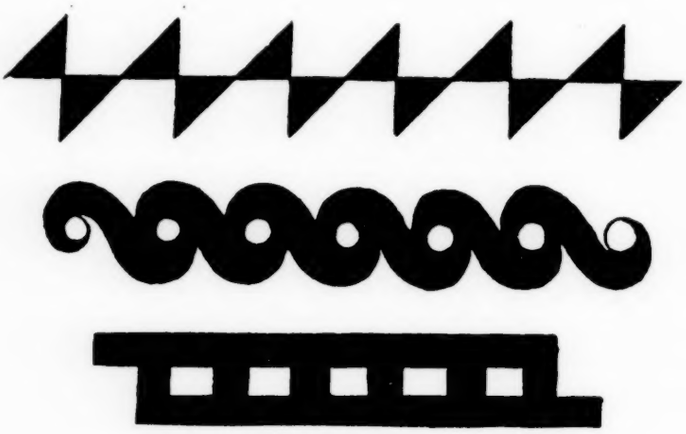


PLATE I. Basic designs (six motive)

are treated as 2 groups—called Series I and Series II—because after the 21st O had made her judgments a hitherto unnoted unevenness in the laying on of the colors in the blue and yellow and green and red was remedied. This change gave a little less value contrast in the B-Y and a little more in the G-R.

In Part II of the experiment we undertook to secure not only judgments of fitness of different color-combinations to given designs, but also to determine the effect of different rhythms—duple and triple—upon judgments. When we had constructed the series of designs for Part I we had accepted a 6-member design as sufficiently long to show rhythm, but after making a few observations on this design drawn in triple rhythm we concluded that it was not long enought to show a triple rhythm to good advantage. We accordingly decided to use a 9-member design. As this was so much larger than the one used in Part I, it seemed to us that the conditions of the two experiments might not be sufficiently constant for purposes of comparison and accordingly we made another series of designs in duple rhythm, this time of 8 members (this being much more comparable in size to the 9-member design) and presented both the duple rhythm and the triple rhythm series to our Os.[4] To draw these designs larger sheets of paper than had been used in Part I were required but unfortunately paper of the same texture was not procurable. This change in paper somewhat altered the appearance of the colors, more in some colors than in others. In the blue and green both colors were darker and the green more saturated, giving the design as a whole a quite changed appearance; in the blue and yellow the blue being darker gave a greater contrast to the yellow, which in itself remained very much the same. In the black and red, the red was more saturated, giving a sharper contrast with the black which in itself was much the same. In the green and yellow, although the green was darker and more saturated while the yellow remained the same, the design as a whole did not appear very much changed. In the black and yellow the change was negligible. In the green and red, both colors were more saturated and the green darker and in value equal to the red, giving the design as a whole a more sharp-cut appearance. Any effect these changes seem to have had will be fully considered in our discussion of results. Equated with Munsell, the colors were: Black 3; Red 5/12; Blue 5/6; Green 6/6; Yellow 8/10

[4] Factually the 6- and 8-member designs showed the same tendencies.

The size of the paper on which the designs were drawn was 6 3/4 in. × 10 in. The procedure in Part II was that followed in Part I, except that here judgments were made on 2 series of designs. Judgments on the duple rhythm were given first, and after a 2-minute rest period judgments on the triple. There were 30 Os for this part of the experiment, 8 of whom had also given judgments in Part I. Table I shows the results of these first 2 parts of the experiment.

TABLE I

JUDGMENTS OF Os (IN PERCENTAGES) ON FITNESS OF COLOR-COMBINATIONS TO LINE DESIGNS, BOTH IN DUPLE AND IN TRIPLE RHYTHM

Bl-R (Black-Red); B-G (Blue-Green); B-Y (Blue-Yellow); G-R (Green-Red); Bl-Y (Black-Yellow); G-Y (Green-Yellow). DC = Design Choice; CC = Color Choice. R = Rectangular; C = Curving; A = Sharp-Angled. *Duple Rhythm:* Series 1 (S1), 21 Os; Series 2 (S2), 41 Os; S3, 30 Os. *Triple Rhythm:* Series 4 (S4), 30 Os. In DC O chose for a given color combination 1 of the 3 designs; in CC for a given design 1 of the 6 color combinations.

	Bl-R		B-G		B-Y		G-R		Bl-Y		G-Y	
	DC %	CC %	DC %	CC %	DC %	CC %	DC %	CC %	DC %	CC %	DC %	CC %
S1												
R	23.8	19.0	23.8	4.7	23.8	00.0	42.8	00.0	57.1	76.1	28.5	00.0
C	23.8	9.5	66.6	42.8	52.3	28.5	42.8	4.7	14.2	00.0	61.9	14.2
A	52.3	80.9	9.5	4.7	23.8	00.0	14.2	00.0	28.5	14.2	9.5	00.0
S2												
R	53.6	39.0	7.3	00.0	24.3	00.0	29.2	9.7	53.6	48.7	26.8	2.4
C	12.2	7.3	82.9	39.0	60.9	31.7	24.3	7.3	14.6	7.3	68.2	7.3
A	34.1	53.6	9.7	7.3	14.6	2.4	46.3	7.3	31.7	29.2	4.8	00.0
S3												
R	50.0	50.0	6.6	00.0	20.0	6.6	23.3	10.0	53.3	33.3	13.3	00.0
C	13.3	13.3	73.3	46.6	56.6	3.3	26.6	00.0	16.6	16.6	76.6	20.0
A	36.6	46.6	20.0	6.6	23.3	6.6	50.0	20.0	30.0	16.6	10.0	3.3
S4												
R	23.3	16.6	13.3	3.3	36.6	3.3	53.3	26.6	46.6	46.6	26.6	3.3
C	40.0	20.0	63.3	33.3	43.3	16.6	36.6	10.0	20.0	10.0	70.0	10.0
A	36.6	50.0	23.3	3.3	20.0	6.6	10.0	13.3	33.3	20.0	3.3	6.6

From Table I, we draw for the Color Choice, duple rhythm, the following conclusions. Bl-R is always more, generally much more, fitting [5] to angles than is any other color combination, and angles are in each of the 3 series much higher [5] than curves, and in 2 series higher [5] than rectangles with which in the third they are about equal. Rectangles are always much higher than curves. Bl-Y is in 2 series judged more fitting [5] to rectangles than is any other color combination (Bl-R being higher in the third); and rectangles are in all 3 series much higher than curves or angles. Angles in 2 series are much higher than curves while in the third both are equally low (16.6 percent).

[5] The expressions 'much more fitting' and 'much higher' are used where the percentage is in all averages *at least* twice that with which it is compared; and 'more fitting' and 'higher' where the percentage is higher but not in all averages as much as twice. In most cases, where 'much' is used it can be seen to mean more than twice.

Curves are always low[6] in Bl-R and Bl-Y. B-G is in all 3 series judged more fitting to the curves than is any other color combination, and judged much more fitting to the curves than to either angles or rectangles. B-Y is in two series judged much more fitting to the curves than to rectangles or angles. In series 3, however, the percentage for all 3 designs is low. That B-Y in series 3 is for curves so much lower than in series 1 and 2, we attribute to the greater value contrast we have pointed out existed in series 3. Since G-Y in Color Choice receives for the 3 series in duple an average of 1.9 percent for the rectangles and angles together, and one of 13.0 percent for the curves, no tendency for fitness of G-Y to any of the 3 is indicated. However, here as well as in Design Choice one sees that it is more fitting to curves than to the other 2. G-R is very low for all 3 designs in all 3 series, except that in series 3 it receives 20 percent for angles.

In the Design Choice, duple rhythm, we find very much the same picture as in Color Choice. Bl-R is in all 3 series judged much more fitting to angles than to curves, and in 2 series rectangles are much higher than curves, while in the 3d rectangles and curves receive an equal score—23.8 percent. In the 1st series angles are much higher than rectangles while in series 2 and 3 rectangles are higher than angles. Bl-Y is in all 3 series much higher in rectangles than in curves and in 2 of the series much higher in angles than in curves—in the 3d angles are almost twice the curves. Rectangles are always higher than angles. B-G, B-Y, G-Y, are in all cases judged much more fitting to curves than to either angles or rectangles. G-R has in series 1 for angles only 1/3 the score of curves or rectangles, but is in the neighborhood of doubling them in series 2 and 3. It should be recalled that the G-R had changed its appearance: in series 2 there was more value contrast than in 1, and in 3 it had a 'more sharp-cut appearance.' These changes may have accounted for the rise in angles in series 2 and 3—as well as for that in Color Choice series 3—and may indicate that G-R when sufficiently contrasted in value may be fitting to angles. Against this, however, is the fact that in Color Choice for angles it is low and that there is, considering both Color and Design Choice, more scatter among the 3 designs for G-R than for the other color combinations.

Considering both the scores in Color and Design Choice, duple rhythm, Bl-Y is shown more fitting to rectangles than to angles. In Bl-Y Design Choice the average of the 3 series for rectangles is 54.6 percent against 30.0 for angles, while in the Color Choice it is 52.7 against 26.6. Bl-R—if one give more significance to Color than to Design Choice, a proceeding that seems justified in that it is a less

[6] 'Low' means 20 percent or less.

forced choice—seems more fitting to angles than to rectangles. In Bl-R Design Choice the average of the 3 series for rectangles is 42.4 against 41.0 for angles, while in the Color Choice it is 36.0 against 60.3. However, in 2 series, Design Choice, rectangles are higher in Bl-R than are angles. Moreover in 1 series (3) in Color Choice rectangles are higher in Bl-R than in Bl-Y. Considering all of the data, then, one is not entitled to say that angles in Bl-R are more fitting than are rectangles. Nor can one say which of Bl-R or Bl-Y is more fitting to rectangles.

Turning to the results in triple rhythm, we find the following agreements with those for duple. In Color Choice curves are higher in B-G than in any other color combination, and in Design Choice are judged much more fitting than either rectangles or angles. G-Y is in Design Choice much higher for curves than for rectangles or angles; and is also higher in Color Choice, although the choice of G-Y is low for all 3 designs. In both Color and Design Choice, rectangles are much higher and angles higher in Bl-Y than are curves, and rectangles are higher than angles.

But we also find in triple rhythm certain changes from the tendencies revealed in the 3 series in duple rhythm. In Bl-R, Design Choice, curves are considerably higher than rectangles and a little higher than angles, while in Color Choice curves are a little higher than rectangles but very much lower than angles. If, as our results in duple rhythm seem to show, Bl-R is considered more fitting to rectangles and to angles than to curves, then this result of series 4 triple rhythm might indicate either that Bl-R in triple rhythm is considered fitting to curves, or that because it is unfitting to rectangles—as both Design and Color Choice suggest it to be—the score for curves rises. Against this 2d alternative, however, is the fact that in all 3 series in duple rhythm angles are judged more fitting to Bl-R than are curves and consequently we should expect that if rectangles are not fitting then angles would receive a higher score—but not curves. In Color Choice angles *do* receive a much higher score—50 percent as against 20 percent. Further consideration will be given to this later. In Design Choice—but not in Color—although curves are still the highest, rectangles in B-Y are much higher relatively to curves than they are in any of the 3 duple series. Here again this might be due to fitness of B-Y, triple rhythm, to rectangles or its unfitness to curves, or merely to contrast effect. Something of a case may be made out for this last alternative when it is recalled that in series 3, duple rhythm, where the value contrast is greater than in series 1 and 2, the score for curves in B-Y drops greatly in comparison to series 1 and 2. Moreover in Color Choice, triple rhythm, curves in B-Y are considerably lower than they are in duple

series 1 and 2 where the contrast is less. It might well be, then, that the B-Y in series 4, triple rhythm, seems positively unfitting to curves because of value contrast and hence that the rectangles are higher. But against this is the fact that in series 3, Design Choice, duple rhythm, curves are judged highly fitting. Certainly in this case the contrast of B-Y had not worked strongly enough against curves to force a choice of rectangles. But on the other hand we must remember that the Design Choice is always a more forced judgment than is Color Choice, and therefore the Color Choice may be more significant. In both Color and Design Choice, angles in G-R are low. This agrees with series 1 but not with series 2 and 3 Design Choice. In Design Choice rectangles are higher than angles or curves, while this is not true in series 2 and 3. In Color Choice rectangles are somewhat higher than in series 1–3. Thus the results for G-R triple Rhythm add weight to our conclusion that no positive tendency for G-R is revealed unless, giving more weight to Color Choice, we say that G-R is shown unfitting to curves.

In Part III in order to check on some of the tendencies revealed by the 4 series of Parts I and II, the following group experiment was performed:

Seventeen designs from those used in series 3 and 4 were numbered and presented in certain pairs to 2 groups of 34 and 30 new Os respectively. As the colors in the designs were changed in appearance when projected on a screen, the following procedure—far from ideal—was followed. Os were seated in rows of chairs placed so that the experimenters could pass down the aisles holding 2 designs in such a way that Os on either side of the aisle could see them. At the beginning of the experiment the following instructions were read: "You will be asked to make 11 judgments. In each case you will be shown two designs. You are to choose between them for fitness of color combination and design. Do not choose on the basis of your liking for the colors or designs, but for the fitness of the color combination and the design. Write down the number of the design which you think better."

Table II shows the pairs of designs in the order in which they were presented, together with the results of the judgments made by the Os. These particular pairs of designs were selected in order in certain cases to verify positive tendencies that had been revealed, and in others to throw light on inconsistencies in results or to clear up certain problems that they had raised. We did not attempt to gain data on all the problems as we felt—considering the method used—that the series of judgments should be short.

The results of pairs 1, 2, 5, 6, 10 throw light upon the problem, raised in our previous discussion, as to whether Bl-R triple rhythm is judged not fitting to rectangles while fitting to curves. In pair 1 Bl-R rectangles in duple are more than 4 times higher than in triple. In pair 6 Bl-R curves in triple rhythm are nearly 3 times higher than Bl-R rectangles, a fact which might seem to indicate a fitness in triple rhythm to the curves and an unfitness to rectangles. But that it is rectangles that are not fitting and that hence in pair 6 curves are forced to receive a high score is suggested by the fact

TABLE II

GROUP JUDGMENTS
64 Os

Pairs	Colors	Design	Rhythm	Percent
1	Bl-R	Rectangles	Duple	82.8
	Bl-R	Rectangles	Triple	17.2
2	Bl-R	Curves	Duple	78.1
	Bl-R	Curves	Triple	21.8
3	B-G	Angles	Duple	62.5
	B-G	Rectangles	Duple	37.5
4	B-G	Curves	Duple	68.7
	B-G	Curves	Triple	31.2
5	Bl-R	Rectangles	Duple	64.0
	Bl-R	Curves	Duple	35.9
6	Bl-R	Curves	Triple	73.4
	Bl-R	Rectangles	Triple	26.5
7	G-R	Rectangles	Duple	51.5
	G-R	Angles	Duple	48.4
8	Bl-R	Angles	Duple	62.5
	Bl-Y	Angles	Duple	37.5
9	Bl-R	Rectangles	Duple	59.3
	Bl-Y	Rectangles	Duple	40.6
10	Bl-R	Rectangles	Triple	37.5
	Bl-Y	Rectangles	Triple	62.5
11	G-R	Rectangles	Triple	48.4
	G-R	Angles	Triple	51.5

that in pair 2 Bl-R curves in duple are over 3 times as high as in triple, and in pair 5 Bl-R duple rectangles are nearly twice as high as as Bl-R duple curves. This seems to indicate that Bl-R neither in duple nor triple is judged to be in itself highly fitting to curves but that in comparison to rectangles in Bl-R, triple rhythm, it is so judged. That Bl-R is judged not fitting to curves was consistently revealed in series 1, 2, and 3, duple rhythm. Only in series 4, triple rhythm, where there was comparison with rectangles in Bl-R, triple rhythm, was it relatively high. That Bl-R, triple rhythm, is not fitting to rectangles is also suggested by pair 10 where Bl-Y triple is almost twice as high as Bl-R triple. Considering all the above facts, it seems strongly indicated that Bl-R, triple rhythm, is not fitting to rectangles, and also highly probable that it is not fitting to curves.

In pair 9, Bl-R, duple rhythm, is judged more fitting to rectangles than is Bl-Y. This corresponds to the results of series 3 but not to 1 and 2. And supports the conclusion that one cannot say that one

rather than the other is judged more fitting to rectangles. In pair 4, B-G, duple rhythm, receives for curves a much higher score than in triple rhythm. And in each of the duple rhythm series, B-G receives a higher score than in the triple rhythm series. In pair 3, B-G duple is judged more fitting to angles than to rectangles, and series 2 and 3, but not 1 Color Choice had given a slightly higher score to angles. In pair 8, angles in Bl-R duple are higher than in Bl-Y duple. This corresponds with the tendency shown in all 3 duple series. In pair 7, the G-R rectangles in duple rhythm are about equal to the angles; and in pair 11, G-R rectangles in triple are also about equal to angles. This is in accord with the results of Parts I and II, for in none of the 3 series does G-R show a decided tendency. Scatter is present throughout.

General Conclusions and Discussion

In answer to our question as to whether certain color combinations are more fitting to certain designs than are others, we can reply that this is certainly true of the designs and color combinations used in our experiments. And to our question as to whether certain types of design are more fitting to certain types of rhythm than are others, we can reply that we did not find this true except probably in rectangles in Bl-R and possibly in B-Y.

Bl-R, except in triple rhythm and in series 1, Design Choice, is much more fitting to the straight line designs—angles and rectangles—than to curves. That Bl-R, triple rhythm, was higher for curves than for rectangles we argued—from results of the group experiment—to be due to its unfitness to rectangles rather than its fitness to curves. As to why in Design Choice, triple rhythm, curves and angles were about equal, we have no data to answer. But we noted that in Color Choice, probably the more significant choice, angles were much higher (2 1/2 times) than curves. Nor have we data to say why in series 1, Design Choice, rectangles and angles are equal, 23.8 percent. Bl-Y is in both duple and triple rhythm much more fitting to rectangles than to curves. Angles are more fitting than curves except in Color Choice, series 3, where both curves and angles are low. All 4 series show that Bl-Y is more fitting to rectangles than to angles, and in Pair 8 in the group experiment Bl-Y is only about half as fitting to angles as is Bl-R. Moreover, in both series 4 and the group experiment, rectangles in Bl-Y, triple rhythm, are higher than in Bl-R.

These 2 color combinations, Bl-Y and Bl-R, are the only ones which our results show fitting to rectangles; all other combinations are consistently low except B-Y Design Choice, and G-R Design and Color Choice in series 4, triple rhythm. Attempt has already been

made to explain this exception in the case of B-Y, and in regard to G-R it may be said that since Bl-R triple rhythm is judged so unfitting to rectangles, the score for G-R may rise for this reason. Moreover, the choice for G-R is only 26.6 percent.

Bl-R is the only color combination shown fitting to angles—although G-R, when sufficiently contrasted in value, suggests some evidence of fitness, and in series 2 Color Choice angles in Bl-Y are 29.2 percent as against 53.6 in Bl-R.

B-G in both duple and triple rhythm is much more fitting to the curve design than to either of the straight line designs. In curves it is higher than any other color combination, although B-Y in certain cases approaches it. We have already spoken of the difference in judgments in regard to B-Y for curves in Parts I and II. In series 1 and 2 B-Y receives a comparatively high score, both in Design Choice and Color Choice. In series 3 it is high in Design Choice but in Color Choice drops very low—to 3.3 percent, though in series 4 this rose to 16.7 percent. These low scores we attributed to the changed appearance of B in series 3 and 4 and the consequent changed appearance of the combination B-Y. Much greater value contrast was present in series 3 and 4 than in series 1 and 2, and for this reason it may have been less fitting to curves. B-G, contrary to what preliminary experiments had led us to expect, is not judged more suited to triple than to duple rhythm.

In brief summary we can then say that under the conditions of our experiment: (1) Bl-R is for angles judged most fitting. (2) Bl-R and Bl-Y are for rectangles judged most fitting, the evidence being conflicting as to which is more fitting. (3) Neither Bl-R nor Bl-Y is judged fitting to curves, certainly not in duple rhythm and not, we argued, in triple; (4) B-G is judged highly fitting to curves; (5) B-G is judged not fitting to angles and to rectangles; (6) B-Y is sometimes judged fitting to curves, sometimes not. This, we concluded, is probably due to the degree of value contrast present; (7) B-Y is judged not fitting to angles and to rectangles with 1 exception—rectangles in triple rhythm. This inconsistency our results did not enable us to completely explain. (8) G-R shows no consistent fitness or unfitness for any of the 3 types of design although fitness in angles is sometimes suggested; (9) G-Y is judged more fitting to curves than to angles or rectangles, but Color Choice shows it low for all 3 designs.

If, as our results indicate, certain color-combinations are more fitting to certain designs than are others, what is the reason, or reasons, for this fitness? On what basis are the judgments made? Although we have no introspective reports from our Os on which we

might base an answer, we should like to point out and discuss three possible bases for judgment.

1. Judgments of fitness and unfitness may rest upon an immediate liking or disliking. What is liked, is judged fit; what disliked, unfit. But such an answer rests upon the assumption that stimuli, such as those used in our experiment, do as a matter of course arouse states of liking and disliking. This we very much doubt. They may; but judged from our own experience they certainly need not. To the writer it seems that much psychological experimentation on the problems of color and color-combinations and of the expressiveness of these has been vitiated by the assumption that given certain stimuli, states of feeling automatically arise. But the very conditions of most of these experiments work against such arousal. That many stimuli given in rather rapid succession should arouse genuine states of feeling seems to us not to agree with the facts of affective life. It is possible that under certain conditions of physiological well-being and with not too many successive stimuli, different states of feeling can be aroused, but to suppose that these conditions will be fulfilled whenever Os are called upon to make judgments seems to us completely gratuitous and to be contrary to fact. Affective life is a much more selective and independent thing than this. Moreover the amount of agreement of judgment found in this experiment would if based upon liking and disliking indicate a greater degree of agreement among affective judgments than is often admitted. One thing that can be asserted on the basis of our results is that the judgments of fitness did not rest upon color preferences, for if this were true, such shifts as those that occurred in the case of different designs in, *e.g.*, black and red or in blue and green would not be possible. No color-combination was judged high for all the designs. Nor was any single color, whatever it was combined with, so judged. To the writer, then, liking is not a plausible basis for the judgments.

2. The judgments may rest upon similarities of attributive characteristics of the colors and the designs. A number of recent experiments upon intersensory relations suggest such a possibility—but just what in the present case such similarities would be, is difficult to say. Hornbostel [7] equated odors, sounds, greys—contending for brightness as the common element in these stimuli on which equations were made. Usnadze,[8] and Fox,[9] have equated sound complexes and visual designs. Among other characteristics roundness and

[7] Von Hornbostel, Erich, Ueber Geruchshelligkeit, *Pflüg. Arch. ges. Physiol.*, 1931, 227, 517–538.

[8] Usnadze, D., Ein experimentaller Beitrag zur Problem der psychologischen Grundlagen der Namengebung, *Psychol. Forsch.*, 1924, 5, 24–43.

[9] Fox, C. W., An experimental study of naming, *Amer. J. Psychol.*, 1935, 47, 545–580.

sharpness were shown to be especially significant for these equations. But leaving aside consideration of other experiments and turning to our own, let us consider what similarities in color-combinations and in designs we can find. Are these similarities to be looked for in the single colors and lines or only in their combinations? Colors in themselves differ in hue, value, saturation. What is there in lines or line designs comparable to these? Is there, *e.g.*, some way in which a hue, *e.g.*, red or blue or green, or the value (light or dark) or a certain degree of saturation, is like the curve or the straight line? And if these characteristics are not applicable to lines as such, are they to line designs? Other characteristics that belong to single colors are their approach and recession, their warmness and coolness, their hardness and softness. But can these be applied to color-combinations? Red is a warm, approaching, hard color; green a cool, retreating, soft one. Which of these characteristics, if any belong to the combination G-R? And since yellow shares these same characteristics with red, and blue with green, the same question is applicable to the combination B-Y, as it is also to the combination G-Y. Blue and green have the same characteristics as far as the above go. Shall we then apply these characteristics to the combination B-G? Do Bl-R and Bl-Y—the other two combinations which we used in our experiment—have the opposed characteristics present in B-Y, G-R, etc.?

Turning to lines and line designs, can we say that any of the above characteristics apply? Is perhaps a straight line warm, approaching, hard; a curving one cool, retreating, soft? And the same question in regard to designs. Curves *are* often spoken of as soft, angles as hard. But is this meant in other than a figurative sense? It will be said, perhaps, that the latter statement is no more figurative than is the one which attributes warmth and approach to red and coolness and retreat to green. But as we understand it, some at least of those who attribute to red and green these characteristics mean it in a literal, not a figurative, sense.

When we come to the characteristics which are generally accepted as belonging to colors in combination, contrast seems foremost in importance. In dealing with color contrast we must consider value, hue, and saturation. In the particular combinations used in our experiment hue-contrast is represented by B-Y and by G-R; high value-contrast by the Bl-Y, medium by the B-Y in series 3, low by all the other combinations. It is true that we have spoken of Bl-R as strongly contrasting, but since our black was low in value and red only medium in value, the contrast we referred to could not be a value-contrast. Was it a saturation contrast? Hardly, since black is supposed not to have saturation. Nevertheless one should note

how highly saturated our red was—almost as red a red as the Munsell system presents. And on the other hand our black in no sense appeared as black as black could be. Are there perhaps value steps at which black, grey, white appear at their maximum—just as colors at certain values attain their greatest saturation? At any rate we would argue that in some way black and red do contrast, just as every-day terminology accepts. As to the saturation contrast in our other color combinations it was high [10] in G-R, equally so in both 6 and 9 member series. However, it should be noted that green and red are both more saturated in series 3 and 4 than in 1 and 2. This greater saturation may, rather than the value-contrast we attributed it to, account for certain results in regard to G-R. B-Y had medium saturation contrast, equally so in all series, while B-G had low in series 1 and 2 and none in 3 and 4. G-Y was high in series 1 and 2, medium in 3 and 4. Bl-Y, like Bl-R, do not properly speaking, have saturation contrast—yet do not they also contrast in some way other than their value contrast? Contrast, then, does legitimately belong to colors in combination, but is there any meaning in saying it belongs to designs? Our results do show that Bl-R and Bl-Y are fitting to straight line designs and not fitting to curves, while on the other hand B-G is fitting to curves but not to straight line designs. Now our blue and green had low contrast value in hue, saturation, value; while black and yellow had at least strong *value* contrast. And in some way black and red seemed to contrast. Is it perhaps legitimate to say that our sharp-angled and rectangular (right angled) designs do have contrast in that lines of marked different directions go to make them up, while the curve design does not have such contrast in that the lines seem rather to flow gradually into each other? If contrast could thus be applied to the designs, a basis of similarity between color-combinations and designs would have been found. Some results of our experiments, however, speak against similarity of contrast as the basis for the judgments. B-Y, next to B-G, is fitting to curves. This makes a color combination of hue contrast fitting to curves. On the other hand B-Y is never more than medium in both value and saturation contrast and in series 3 and 4 where the value contrast differed by 3 steps while in series 1 and 2 by only 1, the score for B-Y dropped radically. This might suggest that lack of value contrast rather than color combination made B-Y suited to curves in those cases where it was so judged. But that low value contrast alone is not at the basis of all the judgments of fitness for curves, and high value contrast for straight line designs is shown by the fact that Bl-R, low in value-contrast, is not judged suited to curves but

[10] High is used where there is a difference of 4 steps or more according to Munsell, medium where 2 or 3 steps, low where 1.

is to straightline designs. On the other hand G-Y, low (1 step difference) in value-contrast in series 1 and 2 and medium (2 steps) in 3 and 4 is more suited to curves than to straight line designs. Yet G-R, medium in value contrast in series 1 and 2 and low in 3 and 4, is not judged suited to curves and not unequivocally to straight line designs.

Nor will saturation contrast alone explain the results, for although it is low in B-G, judged most suited to curves, yet G-Y, high in series 1 and 2 and medium in 3 and 4 is judged more fitting to curves than to straight line designs. And G-R, high in all 4 series, is yet not consistently judged fitting to straight line designs. Considering all of the above facts it seems clear to the writer that contrast alone in one or in several of its dimensions cannot alone account for our results. Color-combination, as such, plays a role. Not that we would hold that a given color combination, whatever the value and saturation of its component members, would always be judged fitting to a given design; we have no evidence to support this, nor should we expect it to prove true. Moreover, we have already presented possible evidence against this in the case of B-Y for curves. But on the basis of our results we do hold that colors of given saturation and value will when combined be judged fitting to certain designs. Thus fitness comes not from similarity in one particular element but from the whole which both line designs and colors in combination are. Apropos of this point it may be well to analyze a little more fully our results from the angle of whether it is combination of colors rather than single colors as such that determines our results. If it is a single color rather than its combination with another given color that determines the fitness of this combination to a design, then in whatever combination this color occurs we should expect the combination to be fitting to the design. Our results do not prove this to be true. Bl-Y, *e.g.*, is fitting to angles and rectangles but B-Y is not. It cannot, then, be yellow alone that determines the fitness of Bl-Y. Likewise Bl-R is fitting to angles and rectangles but G-R is not. Red alone then cannot determine this fitness. B-G is fitting to curves but G-R is not. Greenness as such cannot be the basis of this fitness. Something then in the combination rather than in the separate colors as such seems the basis of the judgments of fitness.

Harmony is another characteristic which is attributed to colors in combination. But how this would be applicable to the line designs is difficult to see. Incidentally it may be pointed out that harmony of color-combination cannot be the basis of judgment of fitness of color-combination and design since colors judged fit for one design are not so judged for others. Other than contrast and harmony no characteristics known to be applicable to colors in combina-

tion occur to the writer as ones which might also be applied to designs and thus furnish similarity of attributive characteristics on which the judgments of fitness might rest. Although, as the above discussion has suggested, there may be certain attributive characteristics which color-combinations and designs have in common that may afford basis for judgments of their fitness to go together, the writer sees as an even more probable basis the third possibility now to be discussed.

3. The judgments of fitness may rest upon similarities in expressive characters. What specifically are the expressive values of different lines, colors, color-combinations, is a question of much disagreement and debate. This is not the place to discuss these, but a good resumé of the experiments which have been performed upon this problem may be found in Chapters 4 and 6 of Chandler's 'Beauty and Human Nature,"[11] and a reading of these chapters will show how much in need of further experimentation the problem is. The writer has such experiments under way and from data so far collected she believes that similarities in expressive value afford a basis for such judgments of fitness as have been revealed by the experiments reported in this article. Colors, color-combinations, lines, line-designs, have an intrinsic emotional expressiveness which we would contend belong as immediately and as certainly to them as do their qualitative characteristics of hue or of line-form, and this despite the fact that these have not been given specific names and that imposed associative habits have dulled their recognition. To Bl-R belongs an intrinsic emotional expressiveness similar to that which belongs to a straight line design. And it may be remarked here that the contrast value we have attributed to Bl-R may rather be an expressive characteristic—something that might perhaps be called 'Aggressiveness.' The intrinsic emotional expressiveness of B-G is similar to that of a curving design. Support for these statements will, we believe, be furnished by our experimental work now being carried out. But even without this work the experimental literature affords certain confirmations of the truth of the contention. Lundholm [12] and Poffenberger,[13] *e.g.*, have found the moods quiet and gentle to be expressed by curves, furious and agitating by angles. Bullough [14] has found 'character' aspects of colors as follows: Colors which contain red are affectionate and frank, those containing blue reserved and distant. Yellow is cheerful and sparkling. Green has

[11] Chandler, A. R., *Beauty and Human Nature*, New York: D. Appleton-Century, 1934.

[12] Lundholm, H., The affective tone of lines; experimental researches, *Psychol. Rev.*, 1921, **28**, 43–60.

[13] Poffenberger, A. T., *Psychology in Advertising*, Chap. 15.

[14] Bullough, E., The perceptive problem in the aesthetic appreciation of single colors, *Br. J. Psychol.*, 1908, **2**, 406–463.

a solid 'bourgeois' character—is restful. In purple there is an inclination to morbidness. Stefanescu-Goanga [15] found red to be aggressive while green was quieting, soft, friendly, cheerful. Yellow is stirring, warm, cheerful. Blue is quieting, somewhat depressing, peaceful, serious. Violet is depressing, yearning, sad, very melancholy. Von Allesch,[16] although finding expressions varying according to different conditioning factors, yet insists that these are not merely extraneous associations but are perceived as in the colors. He also found that expressiveness belongs to color-combinations but that this expressiveness is often not that belonging to the colors considered separately. All in all, then, there is considerable evidence that expressiveness is a characteristic which belongs to colors, color-combinations, lines. Moreover from the results of the experimental work referred to above certain applications to our results may be made, *e.g.*, Stefanescu-Goanga found red to be aggressive; green, and blue, quieting. Lundholm and Poffenberger found quiet and gentle to be expressed by curves, furious and agitating by angles. We found that neither of our combinations with red was judged suited to curves while red and black was to angles. Blue and green was judged suited to curves but not to angles.

And thus although it is clear that much more experimental work is necessary before it will be established just what is the specific expressiveness belonging to different colors, lines, etc., enough has been done to show that such specificity of expressiveness is a fact. And to the writer the similarity of expressiveness is the most plausible explanation of the agreements in judgments of fitness revealed by the experiments here reported. As argued above, both liking and attributive characteristics seem to her very improbable as bases of such judgments.

(Manuscript received July 30, 1941)

[15] Stefanescu-Goanga, F., Experimentelle Untersuchungen zur Gefuehlsbetonung der Farben, *Psychol. Stud.*, 1911, 7, 284–335.

[16] Von Allesch, G. J., Die Aesthetische Erscheinungsweise der Farbe, *Psychol. Forsch.* 1925, 6, 1–91, 215–281.

NON-ALIMENTARY COMPONENTS IN THE FOOD-REINFORCEMENT OF CONDITIONED FORELIMB-FLEXION IN FOOD-SATIATED DOGS

BY W. J. BROGDEN *

Pavlovian Laboratory, Phipps Psychiatric Clinic, Johns Hopkins University School of Medicine

Introduction

Conditioned responses reinforced by food commonly suffer experimental extinction when the food is omitted from the test environment. It is generally assumed that such CR's [1] will show a similar or even more marked decrement when the organism has been satiated by feeding. This assumption is made implicitly by research workers in their experimental procedure, when they test their subjects only after a period of food-deprivation.

In studying conditioned flexion responses in dogs, the author observed samples of rather curious behavior on the part of some of his animals. The conditions of the experiment were such that a biscuit of food was automatically delivered to the animal when a conditioned flexion response was evoked by the CS (3). No food was presented following failure of the CR. Sometimes a CR was made to the CS, but when the food was made available in the food-box, the animal did not eat. This behavior might occur on several successive trials, whereupon the animal would eat the accumulation of biscuits. Or the food might not be eaten at all. Even though the animals were tested only after an eighteen to twenty hour fast, it is quite likely that factors other than the period of food-deprivation might have influenced the hunger drive so that the animal failed to eat the food when it was presented. The point of interest in this behavior is not that the animal failed to eat, *but that it made a conditioned response, and then failed to eat.* Such a performance suggests that there might be non-alimentary aspects of the presentation of food, which serve to provide adequate reinforcement of the CR.

The present experiment was designed to test whether there are non-alimentary components which might serve to reinforce a CR, when the alimentary reinforcing component has been eliminated from the experimental situation. The question proposed for experimental

* Department of Psychology, University of Wisconsin.

[1] Abbreviations: US = Unconditioned (original) Stimulus; UR = Unconditioned (original) Response; CS = Conditioned (substitute) Stimulus; and CR = Conditioned Response.

interrogation was as follows: will there be any difference in the frequency of CR between two groups of food-satiated animals, when conditioned flexion is followed by food in one case, and when no food is given with the other group?

Experimental Procedure

Twelve mongrel dogs, secured from an unselected laboratory stock, and naive with regard to all conditioning experiments, served as subjects in this investigation. The animals were placed in a stock within a light-shielded, sound-proofed test-chamber with a one-way glass screen through which the experimenter made his observations. Left forelimb flexion, evoked by shock to the forepaw, was conditioned to the ringing of a bell. The shock circuit was automatically broken whenever flexion occurred, so that a CR brought about shock-avoidance. The stimuli were electrically controlled by a synchronous motor and suitable relays which made each stimulus-sequence identical. The bell was 2 seconds in duration and was followed immediately by the shock which lasted for 1/10 of a second. Twenty trials per day were given as the test-unit. When 100 percent conditioned flexion-response to the bell was obtained, every CR thereafter was followed by a biscuit of food (3). No food was delivered if the bell failed to evoke a CR. The shock was omitted altogether. These training-sessions were given after an 18–20 hour fast by the subjects. When conditioned flexion reinforced by food reached 100 percent, an additional 4 test-periods were given under the same conditions. Thereupon the animals were divided into three groups of four animals each. As soon as they were placed in the test-chamber, the animals of the experimental group were given as many biscuits of food as they would eat. When they failed to eat food which was placed in their mouths, it was assumed that they were satiated, and the experimental session was run as usual. The bell was presented 20 times. If a CR was elicited by the bell, a biscuit of food fell automatically in the food-box. If no CR was evoked by the bell, no food was delivered. These experimental conditions were maintained for 10 test-periods. With the completion of this procedure, 5 test-periods were given in which the animals were not satiated, but reinforcement by food was as usual. The animals of control group I were satiated in the same manner as the experimental animals just prior to the trials given in 10 successive test-periods. However, the CS was never followed by food, no matter whether a CR occurred or not. This group then served as a control upon the influence of the conditions of food-reinforcement in satiated animals. No further trials were given after the completion of the 10 experimental sessions in which these animals were satiated. The experimental procedure for control group II differed from that of control group I in that no food was given at any time during the test-procedure. These animals were neither satiated nor given food-reinforcement after a CR. Extinction was carried out in the usual manner. The bell was presented alone during test-periods which occurred only after an 18 to 20 hour fast. The results for this group make it possible to determine the effect of satiation upon the course of experimental extinction.

Experimental Results

Satiation of the experimental animals produced a decrement in the frequency of conditioned flexion. However, this decrement was not large. There seemed to be a slight tendency for the decrement in frequency of CR to increase as the number of test-periods increased (Fig. 1). The mean frequency of conditioned flexion for the group was 85.00 percent (σ = 2.31 percent) during the 10 periods before which the animals were satiated. For the 10 test-periods (5 before, 5 after satiation) in which food-reinforcement was as usual, the mean frequency of CR was 92.12 percent (σ = 2.16 percent) (Table I). The difference between these means, 7.12 percent (σ = 3.15 percent), when checked by the test devised by Fisher (8),

is a reliable one. Such a difference might be expected once in 100 samples drawn at random from a homogeneous population.

Comparison of the scores of the animals of control group I, satiated but without food-reinforcement, with those of the experimental group, satiated but with food-reinforcement, will reveal whether food-reinforcement in the food-satiated animal is the vari-

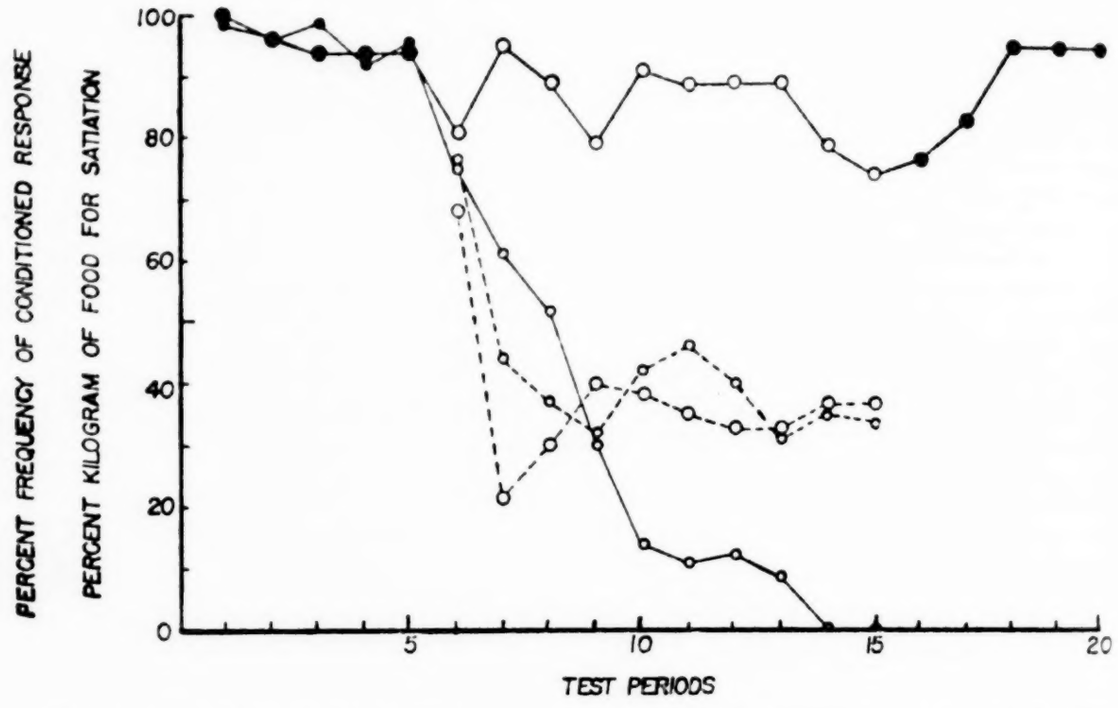


FIG. 1. The large circles in all cases represent data of the experimental animals. The small circles represent data obtained from the animals of control group I. Those circles connected by solid lines represent the mean frequency of CR for the group during a given test-period. Those circles connected by broken lines represent the mean quantity of food in percent kilogram eaten by the group at the beginning of the test-period to reach satiation. The filled circles represent the mean frequency of CR during those test-periods when the animals had been subjected to an 18–20 hour fast. The unfilled circles represent performances during test-periods before which each animal was satiated. During the test-periods when satiation occurred, the experimental animals (large circles) were presented with food following a CR to the CS, but with no food if a CR did not appear. The animals of control group I received no food at any time following the CS during the same period. For further explanation see the text.

able responsible for the high frequency of CR given by the experimental animals. The animals of control group I showed a rapid drop in the frequency of CR (Fig. 1; Table I). The mean frequency of conditioned flexion for the group during the 10 test-periods, when they were satiated but without food-reinforcement during the test-trials, was 26.12 percent (σ = 5.27 percent). Since the comparable mean for the experimental animals is 85.00 percent (σ = 2.31 percent), the difference between the two is 58.88 percent (σ = 5.76 percent). With a p value of less than .0000, this difference is highly reliable.

The experimental group and control group I differed in their performance only when there was a difference in the experimental

TABLE I

The CR scores for each animal are in percent left forelimb flexion response to the 20 presentations of the CS given each test-period. The upper portion of the table gives the frequency of CR under the normal conditions of food-reinforcement. The mean frequency of CR is given for each group. The figure to the right of each mean, in parenthesis, is its standard error. The middle portion of the table gives the frequency of conditioned flexion under changed conditions of reinforcement, as indicated by the heading for each group. Means and their standard errors for the frequency of response are given in percent for the experimental group and control group I. The mean number of test-periods required to reach zero response are given for control groups I and II. The third portion of the table gives data for the experimental group similar to the first section. The first mean is for the last five test-periods, 6–10. The second mean gives the mean frequency of CR for all 10 test-periods, both those in the first and third sections of the table. The abbreviation T-P represents test-period.

T-P	Experimental Group				Control Group I				Control Group II			
	1	2	3	4	5	6	7	8	9	10	11	12
1	100	100	100	100	100	100	100	100	100	100	100	100
2	95	100	90	100	100	90	85	95	100	85	85	100
3	95	90	95	95	100	95	95	100	100	80	95	100
4	95	90	90	100	95	90	95	100	90	90	95	100
5	90	95	95	95	95	95	90	90	100	90	75	95
Mean	95.50% (1.16%)				96.00% (0.80%)				94.00% (1.72%)			
	Satiated Food after CR				Satiated No Food after CR				Unsatiated No Food after CR			
1	65	80	95	85	80	85	90	45	100	55	10	85
2	85	100	95	100	10	90	80	55	90	40	40	80
3	90	100	75	90	65	65	55	25	80	25	65	65
4	75	75	80	85	40	35	40	5	60	10	30	50
5	100	85	95	85	0	25	30	0	45	15	10	70
6	70	95	95	95	0	15	20	10	20	0	30	75
7	80	95	90	90	5	5	40	0	0		0	50
8	90	90	100	75	15	0	15	0				45
9	95	35	100	80	0	0	0	0				20
10	95	45	80	70	0	0	0	0				25
11												10
12												0
Mean	85.00% (2.31%)				26.12% (5.27%)							
Mean					6.75 T-P (1.03 T-P)				8.00 T-P (1.35 T-P)			
6	100	20	85	100								
7	100	60	80	90								
8	90	100	100	90								
9	100	85	100	95								
10	95	85	100	100								
Mean	88.75% (4.26%)											
Mean	92.12% (2.16%)											

conditions, the presence or absence of food-reinforcement of the CR. The frequency of conditioned flexion was similar when food-reinforcement was normal, without satiation. For the 5 periods just prior to the satiation tests, the mean frequency of CR for the experimental group was 95.50 percent (σ = 1.16 percent), and for the animals of control group I, it was 96.00 percent (σ = 0.80 percent). The difference between these two means, 0.50 percent (σ = 1.41 per-

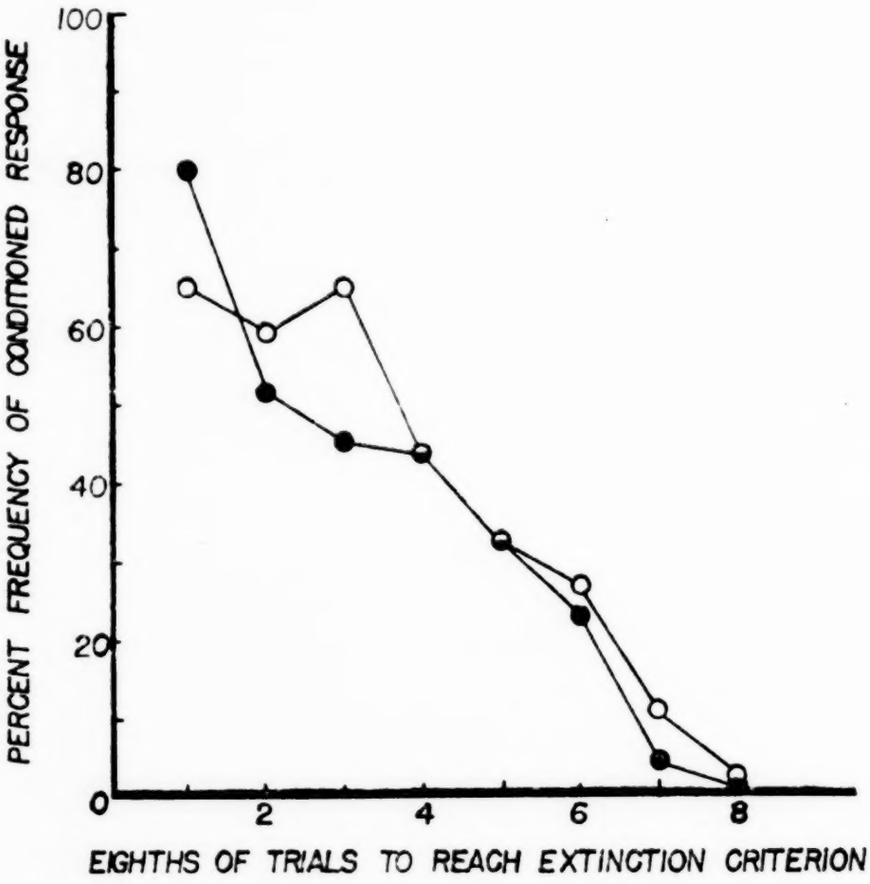


FIG. 2. The filled circles represent the mean frequency of CR for control group I during eighths of the trials required to reach the extinction criterion of 20 CR failures to the 20 presentations of the CS in one test-period. The unfilled circles represent similar data for the animals of control group II. The extinction trials for control group I were given immediately after these animals had been satiated. Those for control group II were given after a period of 18–20 hours of food-deprivation. For further explanation see the text.

cent), with a p value of .39, is without significance. Satiation was equivalent for the two groups, insofar as the amount of food ingested is a reliable and valid measure of satiation. The curves representing the quantity of food eaten to satiation are quite similar (Figure 1). The experimental animals ate a mean of 375 gms. of food (σ = 37 gms.) during the 10 test-periods in which they were satiated; the animals of control group I ate a mean of 418 gms. (σ = 41 gms.) during the same periods. The difference of 43 gms. (σ = 55 gms.), with a p value of .33, is not significant.

The data obtained from the animals of control group II permits a check on those from the animals of control group I. The decrement in frequency of CR, when the CR is not given food-reinforcement, might differ for the satiated animal from that of the hungry animal. Since scores of experimental extinction serve as the norm in assessing the efficacy of different experimental procedures in maintaining a CR, the efficiency of food-reinforcement of the satiated animal in maintaining conditioned flexion might well be inaccurately estimated. Comparison of the two groups shows that the decrement in frequency of CR is quite similar. Zero percent conditioned flexion was reached in a mean of 6.75 test-periods (σ = 1.03 test-periods) by the animals of control group I and in a mean of 8.00 test-periods (σ = 1.35 test-periods) by the animals of control group II. The difference of 1.25 test-periods (σ = 1.71 test-periods), with a p value of .26, is not significant. Vincent curves of the decrement in frequency of CR show that there is little difference in the course of experimental extinction whether the animals have fasted for 18 to 20 hours before the test-periods or whether they have been satiated immediately before being tested (Fig. 2).

Discussion

The marked difference in performance between the animals in the experimental group and those of control group I must have been a function of the conditions of food-presentation, since the presence of this variable in the treatment of the experimental group, and its absence in that of control group I, constituted the only difference in the experimental conditions of the two. Since both groups were completely satiated at the beginning of each test-period, one might assume that the alimentary components of food-presentation were without influence upon the occurrence of conditioned flexion. If such an assumption is made, the obvious non-alimentary components of food-presentation which might have served to reinforce conditioned flexion are the sound of the food release mechanism, the sound of the food falling, and the sight and smell of the food. While there may have been other operations involved in the presentation of food which might have had reinforcing functions, it is difficult to isolate them. Only further experimental work would make it possible to determine those variables, and their relative reinforcing values, which served to maintain conditioned flexion in the food-satiated animal.

In attempting to account for the manner in which these non-alimentary components of food-presentation reinforce conditioned flexion in the food-satiated animal, several alternative hypotheses might be proposed. One might assume that the sound of the food-release mechanism, the sound of food falling, and the sight and smell

of food have a primary reinforcing effect. The likelihood of these variables serving in the conditioning of any response in the satiated animal is very remote.

It is more probable that these factors derived their reinforcing effect from the alimentary component or that they developed as incentives for transformed motives arising out of the conditioning process. That stimuli which occur simultaneously with the primary reinforcing stimulus acquire secondary reinforcing properties has been demonstrated by several authors. Finch and Culler (**7**), Skinner (**9**), and Bugelski (**4**) have shown that an auditory stimulus, which during training accompanied the reinforcing stimulus, prolongs experimental extinction by its presence. The token-reward studies by Wolfe (**11**), Ellson (**6**), Cowles (**5**), Smith (**10**), and others have shown that objects which could later be exchanged for food are effective in the acquisition and maintenance of a number of different responses. The token, however, loses its reinforcing value if it is presented without the subsequent primary reward. The response becomes extinguished and the incentive-value of the token disappears.

Secondary reinforcement seems to depend upon primary reinforcement for its efficacy. Therefore, it could not be the factor responsible for the continued appearance of the CR in the food-satiated animals of the experimental group if satiation had removed the alimentary components of food-presentation. Certainly the evidence of secondary reinforcement, where food is the primary stimulus, points towards its dependence upon those factors which make the alimentary components of food-presentation effective.

If one assumes that during the conditioning process there are new motives developed or old ones transformed, one is faced with the problem of the means whereby this occurs. The view of Woodworth (**12**) that a mechanism (response) may itself become a drive would account for the behavior of the experimental animals only if the control groups exhibited similar behavior. While the hypothesis presented by Allport (**1**) concerning the changing character of motives as they acquire functional autonomy offers one means, the discussion by Anderson (**2**) of the externalization of drives offers the best theoretical structure into which the data of the present experiment might fit. A drive is assumed by Anderson to be some neural, neuro-muscular or neuro-glandular state which provokes persistent behavior on the part of the organism. Some drives are assumed to be dependent for their arousal originally upon conditions within the organism, while others are primarily dependent for their arousal upon the external situation. A drive which was originally elicited by a given set of internal conditions (within the organism), might become aroused by a certain external situation which had been frequently present in

conjunction with the original internal conditions. Through this pairing or 'conditioning,' the external conditions become capable of evoking the drive in the absence of the original internal conditions. The drive, originally dependent upon conditions within the organism, is thus externalized. It is further assumed that the more often the drive is evoked in the company of these external conditions, the more complete is the externalization. Anderson discusses this theory in terms of the hunger drive of the rat in relationship to the learning of a maze habit, and deals with a number of specific implications which follow from the theory.

This theory most adequately accounts for the behavior of the animals in the present experiment. Since Anderson's propositions refer to the maze-learning behavior of the white rat, none of them are so stated that they agree precisely with the various conditions of the present experiment. However, insofar as the events and conditions of the rat maze experiment are analogous to the conditioning experiment, the propositions should hold for both. But because of these differences, no attempt will be made to relate the following discussion to the specific propositions developed by Anderson.

If one assumes that during the process of conditioning, whereby the flexion response evoked by the bell becomes reinforced and maintained by food, the hunger drive becomes more and more dependent upon the external conditions accompanying it, it should eventually be aroused by the external conditions alone, in the absence of the internal conditions. If one further assumes that arousal of the hunger drive is one of the necessary conditions for establishing and evoking the CR examined in this experiment, change in the frequency of the CR with changes in the internal or external conditions accompanying the drive should serve as a test of the hypothesis of externalization of drive.

If externalization of the hunger drive does take place during conditioning, removal of the internal conditions by satiation should produce little disruption of the CR as long as the external conditions (the sound of the food-release mechanism, the sound of food falling, and the sight and smell of food) are present. If one compares the performance of the animals of the experimental group when they were unsatiated with that when they were satiated, it is found that satiation produces a significant, but small, decrement in the frequency of conditioned flexion when food-reinforcement is as usual, even though the food was not eaten.

Removal of the internal conditions of the drive plus the removal of the external conditions accompanying the food-reward, if the drive has become externalized, should result in a greater disruption of conditioning than satiation alone. The animals of control group I, who

were satiated but given no food-reinforcement following a CR, showed a significantly greater decrement in the frequency of CR than did the experimental animals, who were satiated but rewarded with food following a CR.

On the other hand, removal of the external conditions plus satiation produces no greater disruption of the CR than does removal of the external conditions alone. The animals of control group II, unsatiated and without food-reinforcement, gave a decrement in the frequency of CR similar to that given by the animals of control group I, who were satiated but given no food-reinforcement.

Secondary reinforcement was considered inadequate to account for the maintenance of conditioned flexion by the experimental animals when they were satiated, but still presented with food after a CR, because of the dependence of the secondary stimulus for its reinforcing function upon the primary stimulus. Since no studies have been reported which demonstrate that the secondary stimulus retains its reinforcing function in some degree when the internal conditions of a drive, such as the hunger drive, have been removed by satiation, there is no direct evidence that the drive is being externalized concomitantly with or as an integral part of the development of secondary reinforcement. It may be that secondary reinforcement is no more than the externalization of a drive, whereby the arousal of the drive becomes more and more dependent upon the external conditions (secondary stimulus) which accompany the internal conditions originally essential in evoking it. In the studies of secondary reinforcement, the drive might not have been sufficiently externalized for the secondary stimulus to maintain the CR in the complete absence of the primary reinforcing stimulus, even though the internal conditions accompanying arousal of the drive were intact. In the present experiment, the progress of externalization of the hunger drive was sufficient so that the animals of the experimental group continued to give CR's to the CS when the alimentary components of the reinforcing situation were removed by satiation, but the other components remained. Whether or not secondary reinforcement is involved in this performance would depend upon the frequency with which conditioned flexion occurred in a group of hungry animals when the sound of the food-release mechanism and the sound of food falling followed only the elicitation of a CR. If the frequency of CR were greater in this group than that of the animals of control group II, evidence of secondary reinforcement would be obtained. If not, there would be no evidence of secondary reinforcement. Until some such evidence has been obtained, secondary reinforcement can neither be included nor excluded as a possible factor in the present experiment.

Summary of Conclusions

1. The frequency of conditioned flexion reinforced by food showed a small decrement (7.12 percent) when the animals were satiated just prior to the test trials from that which occurred when they had been deprived of food for 18–20 hours prior to the test-period.

2. The animals of a control group that were satiated but not given food-reinforcement after CR's showed complete extinction of conditioned flexion during trials during which the experimental animals responded better than 80 percent of the time.

3. The extinction of conditioning in these animals who were satiated before each test-period was similar to that shown by the animals of a second control group who had been deprived of food for 18–20 hours before the extinction trials were given.

4. The maintenance of conditioning by the satiated animals of the experimental group is considered to be a function of the conditions of food-presentation, since the CR of the animals of control group I, who were satiated but received no food after a CR, was extinguished. That these conditions of food-presentation are non-alimentary in nature is demonstrated by the failure of the experimental animals to eat the food that followed the CR upon every occasion.

5. The means whereby the non-alimentary components of food-presentation acquire a reinforcing function are discussed and the data of the present experiment are considered to support and best fit Anderson's theory of the externalization of drive.

(Manuscript received July 31, 1941)

References

1. Allport, G. W., *Personality*, Henry Holt, New York, 1937, Chap. 7, pp. 190–212.
2. Anderson, E. E., The externalization of drive: I. Theoretical considerations, *Psychol. Rev.*, 1941, **48**, 204–224.
3. Brogden, W. J., Unconditioned stimulus substitution in the conditioning process, *Amer. J. Psychol.*, 1939, **52**, 46–55.
4. Bugelski, R., Extinction with and without sub-goal reinforcement, *J. comp. Psychol.*, 1938, **26**, 121–134.
5. Cowles, J. T., Food-tokens as incentives for learning by chimpanzees, *Comp. Psychol. Monogr.*, 1937, **14**, No. 71, 96 pp.
6. Ellson, D. G., The acquisition of a token-reward habit in dogs, *J. comp. Psychol.*, 1937, **24**, 505–522.
7. Finch, G. and E. Culler, Relation of forgetting to experimental extinction, *Amer. J. Psychol.*, 1935, **47**, 656–662.
8. Fisher, R. A., Application of 'Student's' distribution, *Metron*, 1925, **5**, 90–104.
9. Skinner, B. F., *The behavior of organisms; an experimental analysis*, Appleton-Century, New York, 1938, Chap. 3, pp. 106–108.
10. Smith, M. F., The establishment and extinction of the token reward habit in the cat, *J. gen. Psychol.*, 1939, **20**, 475–486.
11. Wolfe, J. B., Effectiveness of token-rewards for chimpanzees, *Comp. Psychol. Monogr.*, 1936, **12**, No. 60, 72 pp.
12. Woodworth, R. S., *Dynamic Psychology*, Columbia University Press, New York, 1918, Chap. 2, pp. 36–43.

THE MNEMONIC FUNCTION OF INTERFERENCE

BY BRADLEY REYNOLDS

Duke University

The theory of forgetting as a function of interference has held a prominent place among mnemonic theories since Dallenbach and Jenkins (2) concluded, as a result of their study of forgetting during sleep and waking, that forgetting is less a decay of past impressions than an interference of the new with the old. Foucault (1) had earlier advanced a similar explanation, and Heine (1) had demonstrated directly, in a prior study to that of Dallenbach and Jenkins, the savings effected by sleep.

McGeoch (6), as a result of his researches in retroactive inhibition (5, 7, 8), has formulated a theory of forgetting based upon interference. He has (5) in addition to defending the transfer theory of retroactive inhibition as against the perseveration theory of Müller and Pilzecker (1) insisted that retroactive inhibition is essentially the same phenomenon as reproductive inhibition (9) and offers evidence (10) that negative transfer is favored by learning to make a new response to an old stimulus.

Easley (3) has advanced the hypothesis that inhibition as exhibited in the learning situation itself is likewise the result of interference. Specifically, the strength of conflicting responses associated with the same stimulus more or less neutralizes the excitatory strength of the successful response. The same explanation has been suggested more recently by Hull and others (4). According to this view forgetting could occur either through the association of an old stimulus with a new response or through an increase in the excitatory strength of the alternate unsuccessful responses in the original situation. The experiment which is reported here was designed to test one implication of this hypothesis.

Following the cessation of learning, inhibitory potential increases to a level where effective excitatory potential, as a function of the difference between excitatory potential and inhibitory potential, is insufficient to raise the level of the correct response above the limen. This serves to induce 'forgetting' of the correct response. Consequently, to lower the level of inhibitory potential to a point where effective excitatory potential is sufficient to raise the level of the correct response above the limen, should effect a recovery of the originally learned response. The present study demonstrates the effect of reducing inhibitory potentials which have increased in the interim following learning.

Apparatus and Procedure

The apparatus consisted of a panel 45 cm. × 30 cm. with two apertures 10 cm. × 10 cm. centered in the panel and 10 cm. apart. Behind each aperture were four colored lights, red, blue, green, yellow, the light from which diffused through a ground glass screen covering the aperture. Centered in the panel at the top was a small pilot light, and mounted on the upper right corner of the panel was a high frequency buzzer. The buzzer and light, together with a shock induced by electrodes leading off a Muenzinger stimulus box and applied to the subject's wrist and upper arm, were used as stimuli in the learning problem. On the table in front of the panel were the subjects' controls, four buttons each of which turned out a particular light and was marked with a square of colored paper matching the light which the button controlled, pasted immediately in front of it. A Standard Electric Company reaction timer was wired into the apparatus.

The apparatus was so wired[1] that the experimenter could control the lights which appeared in the apertures, control the subjects' buttons so that only the correct light could be turned out, select the stimulus appearing with any pair of lights, and set the timer to record only the reaction time for the correct response.

In the learning problem subjects were required to learn successive series consisting of presentations of three stimuli (shock, buzzer and light), each presented with a pair of colored lights, one of which was correct for each stimulus. For example, in one series shock appeared with red and blue, and red was correct. The subject made a correct choice by turning out the red light by pressing the corresponding button. Pressing any of the other three incorrect buttons did not turn out either of the lights appearing, and did not stop the timer. The light pairs could be varied from left to right to prevent the development of positional habits. The stimuli were presented in random order modified so that in every three presentations each stimulus appeared once, and so that the series could be stopped at any point without effecting any great differential in the number of presentations of any one stimulus over any other.

All subjects learned two series. Each series was presented until the subject had learned the series to a criterion of nine correct choices without marked change in reaction time. In the second series, which followed immediately after the learning of the first, for each stimulus the response incorrect in the first series became the correct response and was paired with a new incorrect response.

After learning the second series half the subjects were given a six minute interpolated activity of solving match tasks; the other half learned a third series in which the incorrect response in the second series became the correct response and was paired with a new incorrect response. The mean time consumed by Group II in learning this third series was 5.7 minutes. All subjects then relearned the first series.

The instructions to the subjects as to the nature of the problem to be learned emphasized the necessity for responding as quickly as possible and of responding upon the basis of the two lights which appeared. The groups were equated on the basis of the length of the first learning series and the number of errors in the same series.

The subjects were thirty-two summer school students at Duke University. There were sixteen subjects in each group. Group I learned two series only, and Group II learned three.

The conditions of the experiment are summarized below: *S* indicates shock; *B*, buzzer; *L*, light; *b*, *r*, *g*, *y*, blue, red, green, and yellow lights. The correct light for each pair is enclosed in parentheses.

Series I	Series II	Series III
S (*r*)*b*	*S* (*b*)*g*	*S* (*g*)*y*
B g(*y*)	*B b*(*g*)	*B r*(*b*)
L y(*b*)	*L r*(*y*)	*L g*(*r*)

Group I: Series I	Group II: Series I
Series II	Series II
Match Tasks	Series III
Series I	Series I

[1] The apparatus was wired by Mr. Lloyd H. Beck, of the Department of Psychology, Duke University.

Results

Table I shows the mean number of trials for each series together with the mean number of trials prior to the criterion, *i.e.*, gross trials less nine. It can be seen that the groups are equal on the basis of trials for Series I. Due to the necessity for balancing the groups as the experiment proceeded, it was not possible to balance the groups

TABLE I

Mean Number of Trials for All Series and Mean Number of Trials Prior Criterion for Series I and Test

	Trials								Mean No. Trials Prior Criterion I		Mean No. Trials Prior Test	
	Series I		Series II		Series III		Test Series I					
	M.	S.D.	M.	S.D.	M.	S.D.	M.	S.D.	M.	S.D.	M.	S.D.
Group I	47.9	19.0	33.2	24.9	—	—	31.1	14.3	38.9	16.7	22.1	14.1
Group II	47.6	13.6	26.7	9.7	30.1	13.8	11.3	2.0	38.6	13.6	2.3	2.0

on the basis of trials for Series II. However, the difference here is highly unreliable, the critical ratio based on the standard error being .98. In the relearning of Series I, Group II shows a marked superiority over Group I, 11.3 as against 31.1, which on the basis of trials necessary to reach the criterion is 2.3 as against 22.1. It is interesting to note that the mean number of trials for Group I on the test

TABLE II

Mean Number of Errors for All Series, and Mean Reaction Times for Last Nine Trials for Series I and Test

	Errors								Mean Reaction Time Series I		Mean Reaction Time Test	
	Series I		Series II		Series III		Test Series I					
	M.	S.D.	M.	S.D.	M.	S.D.	M.	S.D.	M.	S.D.	M.	S.D.
Group I	13.3	6.1	11.2	10.7	—	—	6.0	3.6	52.8	16.6	53.6	10.4
Group II	12.7	4.7	7.0	4.2	7.2	4.4	1.2	1.1	57.5	12.8	38.6	13.6

series is only 2.1 trials less than for Series II, and is 1.0 trials more than the mean for Series III for Group II. The means for both groups for Series II, 33.2 and 26.7, for Group II for Series III, 30.1, and for Group I for relearning Series I, 31.1, would seem to indicate that the transfer effect had reached a maximum in Series II and the differences in favor of Group II in the test series cannot be ascribed to that factor.

An examination of Table II shows both groups equal on the basis of mean errors for Series I. Again there is, for the reasons indicated above, a difference in errors for Series II but this difference also is highly unreliable, the critical ratio based on the standard error being 1.4. Here, as in the case of trials, there is a marked difference in favor of Group II on the test series, 1.2 as against 6.0. The mean reaction times for the last nine trials of both groups for Series I and the test series are presented in Table II. There is no essential difference between any of these means.

TABLE III

TRIAL AND ERROR RATIOS, TRIALS PRIOR CRITERION AND MEAN ERRORS FOR TEST SERIES

	Error Ratio		Trial Ratio		Trials Prior Criterion		Mean No. Errors	
	Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.
Group I	.51	.25	.649	.41	22.13	14.1	6.0	3.6
Group II	.10	.028	.055	.05	2.31	2.0	1.2	1.1
Diff.	.41		.594		19.82		4.8	
S.E. Diff.	.06		.100		3.53		.94	

The data presented in Table III were obtained by the following method: ratios of the errors for each subject on the test over the errors on Series I were computed and a mean of these ratios obtained; the same procedure was followed for trials, nine being subtracted from the number of trials for both series before computing the ratios. The mean error ratio for Group I is .51, and for Group II, .10, giving a difference of .41 with the standard error of this difference .06. The mean trial ratio for Group I is .649, for Group II, .055, a difference of .594 with a standard error of the difference of .100. The difference in favor of Group II for trials on the test series prior to the criterion is 19.82 with a standard error of this difference 3.53, and the difference in errors for the test series 4.8 with a standard error of this difference .94.

DISCUSSION

Upon the basis of the above results there is little question but that increasing in Series II the strength of the alternate unsuccessful response of Series I, produced inhibition of the successful response for Series I. Nor is there any doubt that decreasing the strength of the unsuccessful response for Series I by increasing in Series III the strength of the alternate unsuccessful response for Series II, effected a recovery of the 'forgotten' response, in a later return of the original situation, namely that decrease of inhibition effects recovery.

As an alternate hypothesis to the one suggested above, it might be suggested that what occurs is successive inhibitions in time of two responses with subsequent spontaneous recovery of both. In a

later return of a situation involving these two responses, the earlier inhibited response would have experienced a greater recovery than the later inhibited response and would have a greater excitatory strength than the latter. The chief argument against this explanation is that it seems inconsistent with the usual course of forgetting, whereas the occasional recall of a long forgotten response is consistent with the first stated hypothesis.

A second, more serious objection to the originally stated hypothesis can be raised, however. In the first series, for example, learning a red response to shock necessitates an increase in the effective excitatory potential of this response. This effective excitatory potential is conceived to be a function of the difference between excitatory potential and inhibitory potential of the red response. Assuming the inhibitory potential of the red response to be the excitatory potential of the blue response,[2] then learning blue correct for shock in the second series should *decrease* rather than *increase* the inhibitory potential of the red response. Let us assume at the end of the first series R_r to have a strength of 7 and R_b to have a strength of 4, and at the end of the second series, R_b to have a strength of 7 and R_g a strength of 4, then the effective potential becomes 3, and this functioning as inhibitory potential for the red response in the first series would be less than in the originally learned habit. However, this is to ignore the effect of the red response coming forward to sum with the green response. Due to the fact that the red response is not possible on the basis of the lights appearing in the second series, it can be assumed that its inhibitory effect decreases by a given amount from series to series.[3] This would require a greater excitatory potential for the blue response before the effective excitatory potential was sufficient to raise the level of response above the limen. In the same manner the green response when it moves backward can be assumed to decrease by a given amount. In the second series the assumed strengths might be R_r 4, R_b 11, R_g 4, and in the subsequent recurrence of series I, R_r 7, R_b 11 − (R_g 2). In Series III the assumed strengths might be R_r 1, R_b 7, R_y 4, R_g 13.[4] In the subsequent recurrence of Series I, the assumed strengths could be R_r 7, R_b 11 − (R_g 8 − R_y 1)[5] or R_r 7, R_b 4.[6] The green response of course cannot effect any inhibition of

[2] In order to simplify the explanation the inhibitory effects of the other members of the whole habit sequence are not taken into consideration. The importance of these is conceded.

[3] Other factors, such as the verbal elements present in the situation, also operate to limit the inhibitory effect.

[4] A possibility exists that R_r and R_b, although incompatible in Series I, may sum in this situation. If such is the case, it does not reduce the validity of the explanation being attempted.

[5] R_g is assumed to have a strength of 1 here for the reason that its function in this situation is essentially that of R_r in Series III.

[6] The figures assigned above are strictly arbitrary and are based on no assumptions as to the shape of the gradients involved.

the red response due to the fact that it cannot sum with the blue response, as the only other alternate response to the red response possible in the objective situation.

Summary and Conclusions

Two matched groups of sixteen subjects each learned two series in which three stimuli, shock, buzzer, light, appeared with pairs of colored lights, one of which was correct, one incorrect. In the first series shock appeared with red and blue, red correct, buzzer with yellow and green, yellow correct, light with yellow and blue, blue correct. In the second series the incorrect responses in the first series became correct and were paired with new incorrect responses. At the end of the second series, Group I worked match tasks for six minutes, while Group II learned a third series in which the incorrect responses for Series II became correct and were paired with new incorrect responses. Both groups then relearned Series I. All series were learned to a criterion of nine errorless trials with no marked changes in reaction time.

The results show that for Group I, Series I had suffered a large degree of inhibition while Group II had suffered but little inhibition in the same series.

It is held that these results support the theory that holds forgetting to be a function of transfer effects in learning.

(Manuscript received August 27, 1941)

References

1. Britt, S. H., Retroactive inhibition; a review of the literature, *Psychol. Bull.*, 1935, 32, 381–440.
2. Dallenbach, K. M., and Jenkins, J. G., Obliviscence during sleep and waking, *Amer. J. Psychol.*, 1924, 35, 605–612.
3. Easley, H., The curve of forgetting and the distribution of practice, *J. educ. Psychol.*, 1937, 35, 474–478.
4. Hull, C. L., Hovland, C. I., Ross, R. T., Hall, N., Perkins, D. T., and Fitch, F. B., *Mathematico-Deductive Theory of Rote Learning: A Study in Scientific Methodology*, New Haven: Yale University Press, 1940.
5. McGeoch, J. A., The influence of four different interpolated activities upon retention, *J. exp. Psychol.*, 1931, 14, 400–413.
6. McGeoch, J. A., Forgetting and the law of disuse, *Psychol. Rev.*, 1932, 39, 352–370.
7. McGeoch, J. A., and McDonald, W. T., Meaningful relation and retroactive inhibition, *Amer. J. Psychol.*, 1931, 43, 579–588.
8. McGeoch, J. A., The influence of degree of learning upon retroactive inhibition, *Amer. J. Psychol.*, 1929, 41, 252–262.
9. McGeoch, J. A., McKinney, F., and Peters, H. N., Studies in retroactive inhibition: IX. Retroactive inhibition, reproductive inhibition, and reminiscence, *J. exp. Psychol.*, 1937, 20, 131–143.
10. McGeoch, J. A., and McGeoch, G. O., Studies in retroactive inhibition: X. The influence of similarity of meaning between lists of paired associates, *J. exp. Psychol.*, 1937, 21, 320–329.

DEGREE OF EFFORT: II. QUALITY OF WORK AND TIME OF COMPLETION OF PERFORMANCE TESTS

BY G. K. YACORZYNSKI

Northwestern University Medical School

The degree of effort which an individual is expending is determined by the quality of the methods which he uses to complete a task. If an effort is made to use the more difficult methods more time is spent in completing the problem than if inferior methods are used. It follows, therefore, that on appropriate tasks the quality of the work and time taken to complete the tasks should be related directly to the degree of effort expended. In other words, the quality of the work should be superior, and the time taken to complete the tasks should be longer as more effort is expended. It appeared that the object assembly, the picture arrangement, and the block design test originated by Kohs of the Wechsler-Bellevue (3) scale might yield such results since in the manual describing this scale the first two tests are given part score on the basis of quality of performance, and the Bolles and Goldstein (1), Nadel (2), and Wechsler (3) have noted that qualitative differences are present among individuals in their performance on the Kohs block design test.

METHODS

The three tests of the Wechsler-Bellevue intelligence test were administered in the manner directed in the manual, with the exception that no test was discontinued, even if the time limit given in the manual was exceeded, until the subject had either completed the test to his satisfaction, or stated that he was unable to do so.

Four groups of subjects, with 10 subjects in each group, were selected on the basis of the varying degree of effort expended as determined by the open and blocked runway tests described in the preceding study.* The group which used the same tools on the open and blocked runway tests expended, of course, the greatest degree of effort, and the groups which failed to use 1, 2, or 3 tools on the open runway test, but used these tools when the runway was blocked, exerted progressively less amount of effort. Five trials on the open runway test were used to determine the individual's classification, since, as previously pointed out, this number gives the optimum results in classifying the subjects. The number of trials on the blocked runway test were reduced from five to four since it was determined that this procedure does not affect the individual's score.

The subjects consisted of 18 males and 22 females, with an average age of 35.0 years and a range of 12 to 56 years.

---

*This JOURNAL, March, 1942.

Results

Table 1 shows the average number of items completed correctly (the object assembly test consists of three, the picture arrangement of six, and the Kohs block design of seven items), the average time to complete the items which were correct, and the F and chi-square tests to determine if any significant differences are present.

The only significant difference between the groups is the time taken to complete the Kohs block design.[1] Although on this test

TABLE 1

The average number of items completed correctly and the average time taken to complete the correct items of the Wechsler-Bellevue scale. The 0 group of subjects expended the greatest degree of effort since they did not fail to use any of the tools on the open runway test, the −1, −2, and −3 groups show a progressive decrease in the amount of effort expended since they failed to use 1, 2, or 3 of the tools.

Group	Ave. Number Correct Items			Ave. Time to Complete Items		
	Object Assembly	Picture Arrangement	Kohs Block Design	Object Assembly	Picture Arrangement	Kohs Block Design
0	1.8	3.4	5.6	56.9	27.3	77.1
−1	2.1	3.6	5.6	46.7	16.1	48.6
−2	2.1	3.9	5.1	46.6	20.7	42.0
−3	1.4	3.6	3.9	48.7	23.5	27.4
	χ^2 1.78	0.35	3.82	F *	2.74	11.45

* Variance of group means = 237.26, and of error = 870.30.

there appears to be a general decline of the number of figures completed correctly as the individuals expend less effort, neither on this test nor on the others, however, are any significant differences present.[1]

There were also no significant differences between the groups when the tests were scored in the manner directed in the manual. These results can be expected since it was already shown that there is no relationship between general ability and the degree of effort.

Discussion

The only results that, in part, substantiate the expectation that the quality of performance and the time to complete tasks will be related directly to the degree of effort expended, are obtained on the Kohs block design test. On this test, however, only the differences in the time taken to complete the items are significant. The fact that the number of items completed correctly follows the expected trend might mean that if the individuals in the different groups were homogeneous in general ability a significant difference would be ob-

[1] A more painstaking analysis of the quality of performance yielded the same results.

tained. The reason for this statement is that the scores on the block design test yield high correlations with test results for general ability, and since the subjects used in this study were heterogeneous in this respect, being selected from the general population, the dispersion would be increased and the test for significance lowered. It is also possible that since there are only seven items in the test the number is insufficient to yield statistically significant differences. The same argument, of course, could be used for the results of the object assembly and picture arrangement tests, although there is the further possibility that these two tests actually do not involve qualitatively different results, and, therefore, are inadequate for the purpose for which they were used.

It is of interest that Wechsler (3) in discussing the results on the Kohs block design obtained in standardizing the intelligence test on an adult population states, "There is also a difference of attitude and emotional reaction on the part of the subject which gives us valuable information about him. One can often distinguish the hasty and impulsive individual from the deliberate and careful type, a subject who gives up easily or gets disgusted, from the one who persists and keeps on working even after his time is up, and so on of a number of other temperamental traits which manifest themselves not infrequently" (p. 95). These qualitative observations agree in many respects with the definition which has been used in differentiating the individual on the basis of the degree of effort expended.

Summary

The expectation that on appropriate tasks the quality of the work and the time taken to complete the tasks should be related directly to the degree of effort expended, follows from the definition and the preceding study of this behavior characteristic. Of the object assembly, picture arrangement, and Kohs block design tests of the Wechsler-Bellevue scale administered to four groups of subjects expending different degrees of effort, only the time of completing the Kohs block design test yielded the expected results. The quality of the performance, determined by the number of items correctly completed, on this test follows the expected trend which suggests that under appropriate conditions this trend would be significant.

(Manuscript received July 6, 1941)

References

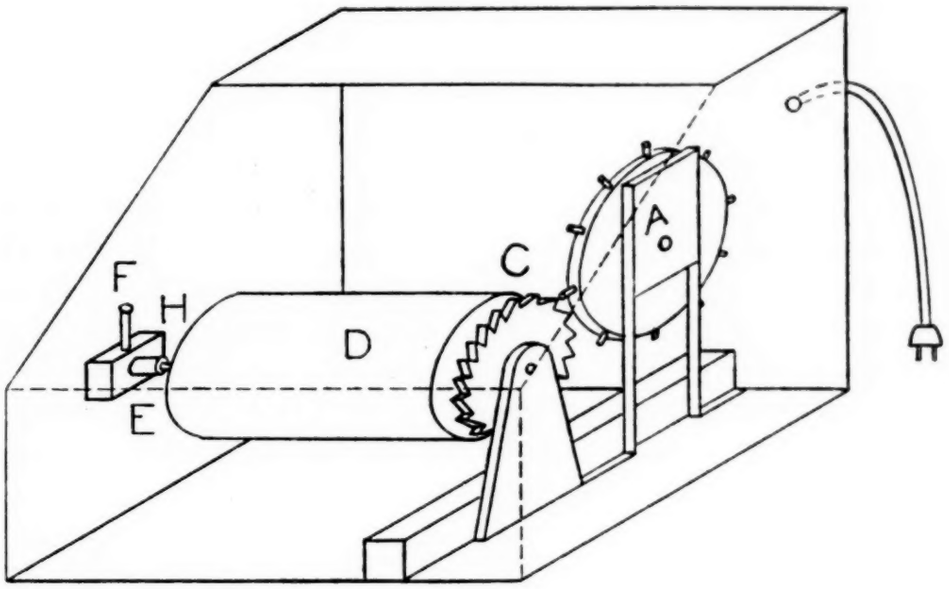
1. Bolles, M. M., and Goldstein, K., A study of the impairment of 'abstract behavior' in schizophrenic patients, *Psychiat. Quart.*, 1938, **12**, 42–65.
2. Nadel, A. B., A qualitative analysis of behavior following cerebral lesions: Diagnosed as primarily affecting the frontal lobes, *Arch. Psychol.*, 1938, **32**, pp. 60.
3. Wechsler, D., *The Measurement of Adult Intelligence*, Baltimore: Williams and Wilkins Co., 1939, pp. 229.

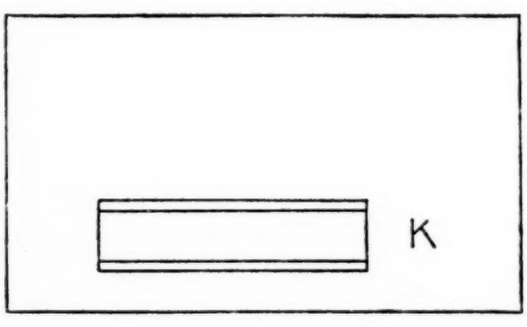
AN INEXPENSIVE, NOISELESS MEMORY APPARATUS

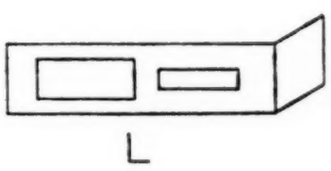
BY JOHN E. WINTER

West Virginia University

The accompanying figures illustrate a simple, effective exposure apparatus that can be assembled for less than $5.00. This particular instrument is operated by a small electric motor such as is used in electric signs, but a motor taken from the cheapest telechron







clock will give perfect satisfaction. The price of the motor need not exceed $2.50. The wheels are made of hard wood, and the drum is an empty paint can.

The motor *A* rotates the sprocket wheel which in turn engages the cog wheel *C*. A metal rod *E* which passes through the drum *D* fits into a socket in the axle of cog wheel *C*. The rod *E* holds the

drum by means of lock nuts and thus the drum travels with cog wheel *C*. The farther end of axle *E* fits into socket *H* which passes through the outer wall and is held in place by lock bolt *F*. Releasing lock bolt *F* makes it possible to remove the drum. Paper can be fastened to the drum by means of rubber bands.

The sprocket wheel has 10 sprockets and the cog wheel has 20 cogs, hence the drum travels just half as fast as the sprocket wheel. The diameter of the drum and cog wheel is 3 1/2 inches, that of the sprocket wheel is 4 1/2 inches. The diameter of the cog wheel is enlarged to enable one sprocket to be completely removed before the next one engages the cog wheel. Since the sprocket wheel is engaged with the cog wheel only at stated intervals, thus making it possible for the cog wheel to move independently in the intervening intervals, the cog wheel is held firmly by means of a spring placed on the axle between the wheel and its support. Thus the drum and cog wheel move only when they are activated by the sprocket wheel, and are at rest in the intervening periods.

The sprocket wheel makes one revolution in 30 seconds, one sprocket thus engaging one cog every 3 seconds. It takes about one second for a sprocket to move the drum to its next position, thus giving a 2-second static exposure for memory material. The adjustable slide *L* moves across aperture *K*, and thus makes possible single or multiple line exposures. The drum makes 20 stops in a single revolution and can thus accommodate up to 20 syllables or other memory units.

(Manuscript received January 26, 1942)